Chapter 13

A SURVEY OF PRACTITIONERS OF EVOLUTIONARY COMPUTATION

Gregory S. Hornby[1] and Tina Yu[2]

[1] *University of California Santa Cruz & NASA Ames Research Center, M/S 269-3, Moffett Field, CA 94035;* [2] *Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada*

Abstract To assist in understanding trends in the field of Evolutionary Computation (EC) and in helping graduates find jobs in EC, we conducted a survey from March 2005 to February 2006 on members of the EC community. The analysis reveals various technology transfer strategies and activities took place during the past 50 years: parallel exploration of multiple application areas; a combination of exploitation and exploration approaches to develop EC applications; and the healthy migration of EC practitioners between different parts of the globe. We believe these emerged and self-organized phenomena contribute to the growth of the field. While there are still challenges in deploying evolutionary computation to industry in a grand scale, the EC community demonstrates the adaptability and resilience necessary to achieve that goal.

1. INTRODUCTION

The field of Evolutionary Computation (EC) has been around for several decades (De Jong, 2006; Fogel, 1998). In recent years, there has been an explosion not only in the different types of biologically inspired algorithms, but also in the number of practitioners in the field. A critical part of this growth and development of the EC field has been the technology transfer of EC from academia to industry and the successful application of EC techniques to real-world problems. To assist in the continued technology transfer of EC techniques from academia to industry, we conducted a survey of EC practitioners working in both academia and industry. This chapter summarizes some of our findings.

The survey was conducted between March 1, 2005 and February 28, 2006 by posting 14 survey questions on the SIGEVO web-site. The survey we ran had three parts. First, it asked several questions about the participant's background.

G.S. Hornby and T. Yu: *A Survey of Practitioners of Evolutionary Computation*, Studies in Computational Intelligence (SCI) **88**, 283–297 (2008)
www.springerlink.com

The second part of the survey had questions on their job information, and the third part was only for non-academic jobs and asked about EC acceptance and applications at that organization. In this survey, we are particularly interested in learning about how participants found their job, how EC techniques are looked on in their organization, which applications they used EC, what problem types that they found EC to be useful and what obstacles they have encountered in applying EC in their organizations.

In this chapter, we report our findings and our observations of the trends in the EC field. Some of the main findings from our results are: there has been an exponential growth in both EC graduates and practitioners; the main source for finding a job has been networking; while most respondents to our survey are in Europe, the most growth of EC in industry has been in North America; the main application areas of EC techniques are multi-objective optimization, classification, data mining and numerical optimization; and the biggest obstacle for the acceptance of EC techniques in industry is that it is poorly understood. In the rest of this chapter, we present the methodology of this survey in Section 2. Section 3 summarizes survey participants' personal information. In Section 4, we report EC-related jobs. EC positions, problem types and application areas are analyzed in Section 5. Section 6 provides data on computer clusters used for EC applications while Section 7 gives the EC acceptance in industry. Suggestions for future surveys are discussed in Section 8. Finally, Section 9 concludes the chapter.

2. METHODOLOGY

The respondents to this survey were not randomly selected but were recruited through a variant of the snowball sampling strategy (Vogt, 1999). The recruiting methods include posting the survey announcement to various EC mailing lists (such as EC-Digest and genetic-programming), e-mailing the announcement to attendants of major EC conferences (such as GECCO-05, GPTP-05, EH-05) and advertising the survey at these conferences. Snowball sampling relies on referrals from the initial subjects to generate other subjects. Although snowball sampling may introduce bias into the study, it can be effective in reaching groups having common characteristics (Atkinson and Flint, 2001). In our case, many EC practitioners are likely to subscribe to EC-related mailing lists and attend EC-related conferences, hence they can be reached by our recruiting approach. However, snowball sampling does not qualify as a random process. Consequently, the results from this survey cannot be generalized to the entire EC practitioner population, regardless of the number of responses received. Nevertheless, these results are still useful for gaining a preliminary picture of EC-practitioners in the world.

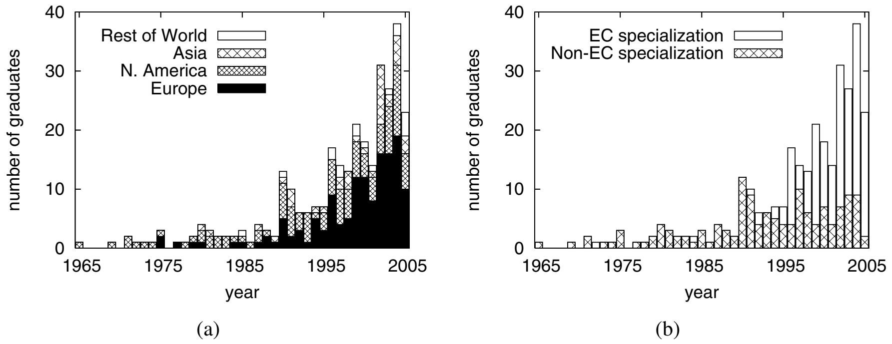


Figure 13-1. This box-plots in this figure break down graduates by: (a) the geographic region in which they studied and (b) whether or not their degree was specialized in EC.

Over the one year time period in which the survey was taken, 324 responses were received, of which 305 had some EC relation, either through graduating with a degree specialized in EC or by using EC in one of their jobs. For the results of this survey, only the 305 responses which had an EC connection were used.

3. PARTICIPANTS INFORMATION

The first part of the survey asked participants to provide their personal information, such as gender, and to answer some questions about the most recent degree that they had received. Of the 305 responses with an EC connection we found that 71.1% of participants have a Ph.D. and the gender split is 87.5% male and 12.5% female. Looking into the geographic regions from which participants graduated, we found that most participants graduated from Europe (46.2%), which is followed by North America (35.7%), Asia (12.5%), Oceania (2.6%), South America (2.0%), and Africa (0.7%). A small number (0.3%) of the survey participants did not answer this question. This data is shown in Figure 13.1(a). When we looked for yearly trends in these percentages based on graduation date, we found that they have remained fairly constant throughout the years.

One change that has occurred over the years is in the amount and the specialization of graduates. As shown in Figure 13.1(b), there is an exponential growth of graduation rate, starting with only a couple of people graduating a year from the 1960's up until the end of the 1980's, at which point the numbers increased dramatically and reach a peak of 36 graduates in 2004. For the first few decades, none of those who graduated in this time period had a degree specialized in EC. The first EC grad does not show up until 1991, and then starting in 1996 the majority of graduates have an EC-specialized degree. This suggests that EC emerged as a field of its own sometime in the mid-1990s.

4. JOB INFORMATION

In total, there were 540 jobs entered by the participants, of which 424 jobs used EC techniques. For determining trends by year, the jobs that were entered by the participants were converted into *yearly positions*. That is, a job from 1997 to 2001 was separated into five positions: one in 1997, one in 1998, and ones in 1999, 2000 and 2001. To limit participants such that they had at most one position in each year, jobs with overlapping years were modified so that the second job started in the year after the first job ended. For example, if a participant had a job from 1995 to 1998 followed by one from 1998 to 2001, the starting year for the second job was changed from 1998 to 1999. Using this method, the 540 jobs were mapped to 3,392 *yearly positions* and the 424 jobs that used EC techniques were mapped to 2,955 EC-related positions.

4.1 Job Sources

After graduation, the next career step for most people is finding a job. By far the most common source of a job was networking, through which 35.1% of participants found a job. This was followed by *other* (25.7%), *supervisor* (17.2%), *postings at university department* (13.8%), *web* (4.7%), *campus career services center* (2.0%) and *mailing list* (0.7%). Looking for differences between those who took a job in academia versus those who took a non-academic job, we found that *networking* was used more for finding a non-academic job (43.0%) than it was for finding one in academia (31.5%). In contrast, the reverse was true for *postings at the university department*: it was used by 16.5% of those who took an academic position but by only 8.0% of those who took a job in industry. Of those who selected *other*, 19 found their position through a listing in a journal or society magazine (such as the Communications of the ACM and IEEE), 13 found their job through an advertisement in the newspaper, 11 founded their own company, and 6 applied and received a research grant.

Further examination of the correlation between the job areas and the job-hunting methods found only a couple of patterns (Figure 13.2(a)). One is that *postings at the university department* helped in finding jobs in Energy, Robotics and Government laboratories, but was of little use for the other job areas. Similarly, the *campus career center* had some success only in finding jobs in Government laboratories and Other. When job-finding methods are analyzed with the job regions, it shows some additional regional trends. *Networking* was used to find over half of the jobs in North America (as well as in Africa, Oceania and South America), but for less than a third of jobs in Europe, and for only 15%

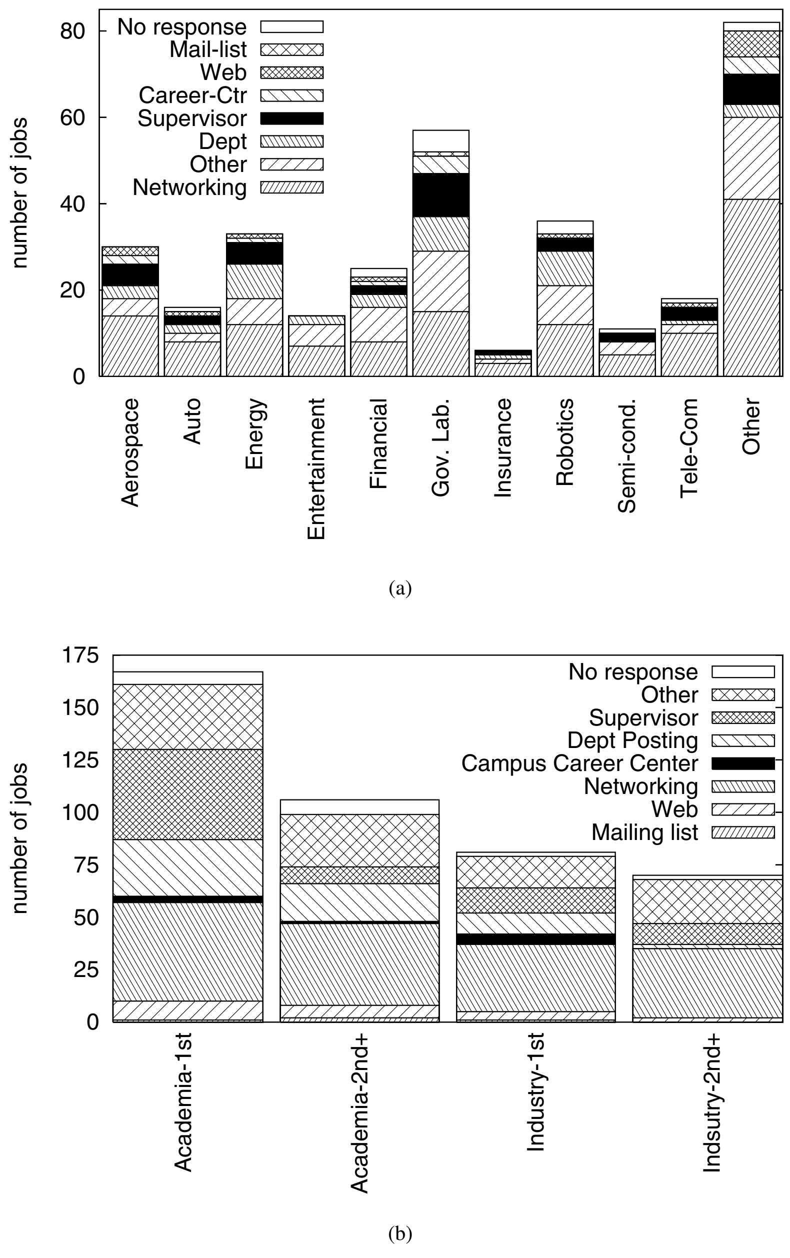


Figure 13-2. This figure contains two graphs which show which job methods were used to find a job by: (a) job area and (b) those going into academia versus those going into industry for their first and later jobs.

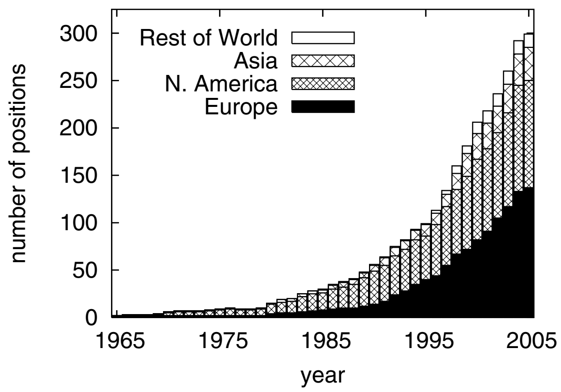


Figure 13-3. A breakdown of EC positions by year and geographic region.

of jobs in Asia. In Europe, *supervisors* helped to find roughly a quarter of all jobs, and they were also helpful in Asia but were not very useful for finding jobs in North America. The *campus career center* was used by a small percentage of the respondents in Asia and North America, but was not used in any other geographic region.

One useful information for job seekers is that different methods were used to find the first jobs and subsequent jobs, according to the survey participants (Figure 13.2(b)). *Supervisors*, not surprisingly, are used a lot more to find the first academic position than to find a second academic position. In contrast, *supervisors* has similar success in helping finding the first job and subsequent jobs in industry. For both academic jobs and jobs in industry, methods specified in the *Other* category were used more frequently for a second job than for the first one. This difference is about 5% for academic positions and about 11% for industry positions, which indicates that the participants used more creative approaches to find their second and later jobs.

4.2 Job Regions

Looking into the distribution of jobs by geographic region, we found that most EC jobs have been in Europe (45%), followed by North America (37%), Asia (10%), Oceania (3%), South America (2%) and then Africa (2%). This geographic distribution of jobs matches closely to the geographic distribution of graduates and suggests a strong correlation between where a graduate studied and where s/he worked. Also, the ratio of positions between the different geographic regions has been fairly constant over the years (Figure 13-3).

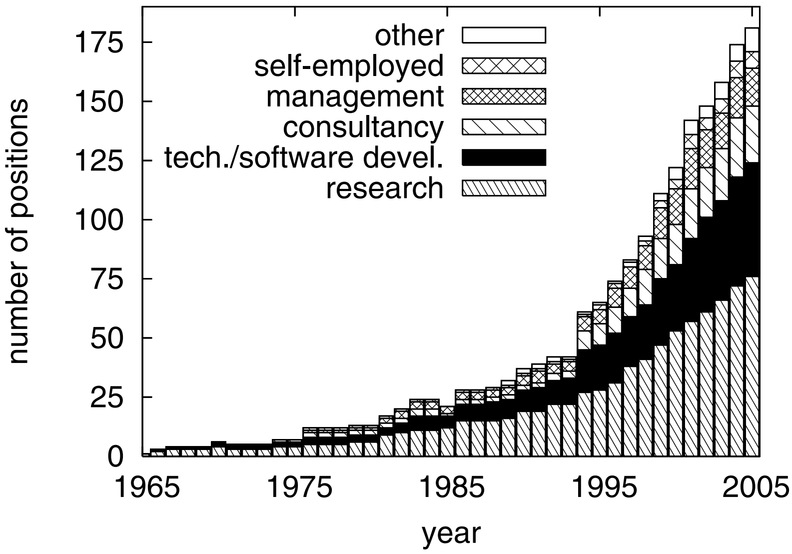
When the job positions are grouped by geographical regions, analyzing the responses over the years reveals that the ratio between positions in industry and in academia has been fairly constant in recent years both in Europe (1:3) and in North America (2:3). In contrast, Asia has experienced a shift in its ratio from being predominantly in industry (100% non-academic in 1981) to being predominantly in academia (more than 75% academic in 2005). For the other geographic regions, the numbers of respondents was too small to give a meaningful interpretation.

Examining the movement of EC graduates for work reveals some interesting trends. First, none of the respondents who graduated with a degree specialized in EC from Africa or South America have left their regions for a job and only 12% of people who graduated in Europe or North America ever move to a different region for work. In contrast, 44% of EC graduates in Asia and 40% of EC graduates in Oceania move at some point after graduation. Second, the direction of movement in Asia, Europe and North America is toward the West. Of those graduates who moved to a different geographic region for a job we found that: 62% of those graduating in Asia moved to Europe at some point, but only 25% ever moved to North America; 70% of those graduating in Europe moved to North America but only 20% moved to Asia; and 67% of those graduating in North America moved to Asia but only 17% moved to Europe for a job. Thirdly, for those people who moved from another region to North America, half moved for jobs in academia and half for jobs in industry, but for those participants who moved to a region other than North America, in all cases they went for academic positions.

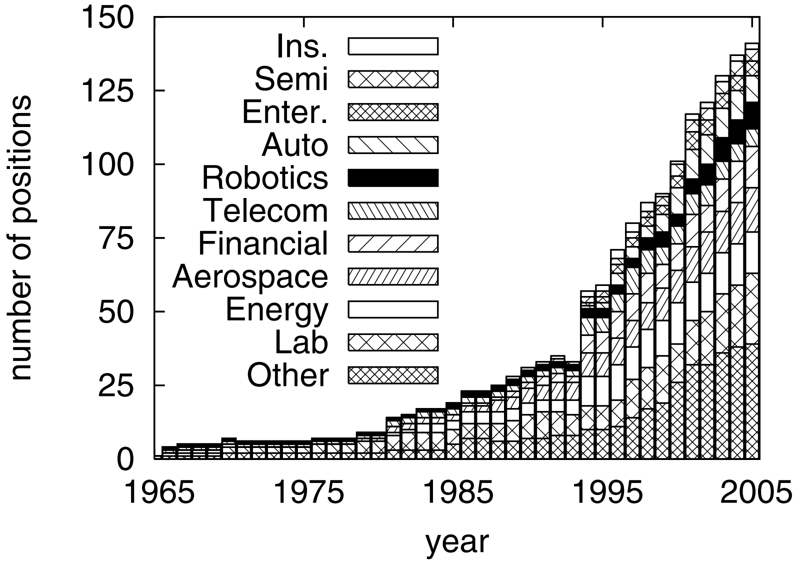
5. EC POSITIONS, PROBLEM TYPES AND APPLICATION AREAS

Once in a job, we are interested in what kind of position in their organization the respondent had, as well as whether or not EC was used and how it was applied.

From our responses we found that there has been an exponential growth in positions in the field, starting with a single EC position in 1965 to just under 300 EC positions in 2005. Breaking this down into academic and non-academic positions, there has been a fairly steady proportion of just under a two-thirds of the positions in a given year being academic and just over a third being non-academic. Figure 13.4(a) is a breakdown of the type of position held for those not working in academia. This figure shows that most industrial EC positions are in research, with a significant number in technical/software development and consultancy. Looking into the types of problems that respondents worked on, we found the following: 40.3% do Multi-objective optimization (MOO), 38.4% do Numerical optimization, 38.0% do Classification, 37.7% do Other, 31.6% do Data mining, 21.2% do Open-ended design, 21.2% do Scheduling, 13.9% do Planning, and 10.1% do Satisfiability/TSP. These values do not add



(a)



(b)

Figure 13-4. This figure contains box plots of: (a) the type of industry position held; and, (b) the application area to which EC is applied.

Table 13-1. Percentage of respondents working in each problem area.

Area	Percentage (%) working in this area	
	Academia	Industry
MOO	38.8	45.9
Classification	38.0	46.6
Num. opt.	36.7	47.0
Other	39.4	34.4
Data mining	28.7	38.6
Scheduling	19.8	37.6
Open-ended design	24.7	24.2
Planning	13.5	21.3
Sat./TSP	10.1	15.6

up to 100% because participants were able to make multiple selections for each job.

For those responses that selected "Other", participants were able to enter a response in a text field. The most popular entries that were given are: optimization and design (24); modeling and simulation (17), EC theory (15); biology and bio-informatics (11); control (11); evolutionary robotics (6); artificial life (5) and neural networks (5). Many of these entries for "Other" fit under the given categories (eg. 'optimization and design' fits under Optimization and/or Open-ended design) with some of the other entries being an application area and not a problem type.

Comparing the distribution of problem-types worked on by academics to that of non-academics, we found a significant difference (see Table 13-1). In general, the percentage of academic positions that are working in a particular problem area is lower than that for non-academic positions. This means that academics tend to focus on fewer problem areas than those outside of academia. Specifically, those participants employed in academic positions average working on 2.24 problem areas whereas those in non-academic positions average working on 2.74 problem areas. Normalizing for this difference, *Scheduling* stands out as the one problem area which is significantly under-investigated by academics as compared to non-academics.

The distribution of problem areas can be further broken down by examining differences by geographic region (Figure 13-5). First, there are more people working on almost every problem area in North America than there are in Europe, even though there are more EC practitioners in Europe than in North America. Since one position can work on multiple problem areas, this indicates that each position in North America tends to work on more problem areas

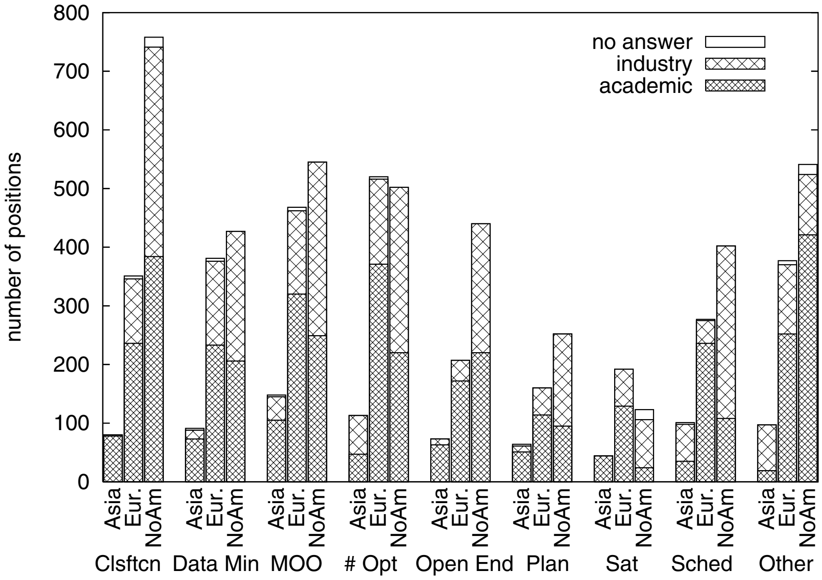


Figure 13-5. A breakdown of EC positions by geographic region, problem area and academia or industry.

than a position in Europe does. In fact, the average number of problems areas worked on by each position for the three different geographic regions is (academic:industry): Asia (2.41:1.92); Europe (2.14:2.39); and North America (2.35:3.30).

Different from the kind of problem being worked on (numerical optimization, scheduling, ...), is the industry to which this problem is being applied (automotive, insurance, ...). Figure 13.4(b) contains a histogram of EC industrial application areas by year.[1] The industry with the largest selection rate is *Other*, which was selected in 37% of all jobs. The most common areas given by those who selected *Other* were: IT (13), consulting (12), biology/medicine related (e.g. Bioinformatics, biomedicine, pharmaceutical) (10), defense and military (7), and various types of engineering (civil, structural or manufacturing) (7). For non-academic jobs, the ways in which EC is reported to be most useful are: *design* (52.3%), *operations* (33.1%), invention (27.8%), *testing* (15.9%) and *other* (14.6%). Of the 31 responses for *other*, 10 were for optimization.

[1] Since each job was allowed to enter multiple application areas the total number of selected application areas can be greater than the number of positions.

Table 13-2. Percentages of job areas that involve work in different problem type.

Job Area	Tot	Percentage working in this problem type.								
	#	Clsf	DM	MOO	NO	Design	Plan	S/TSP	Sched	Oth
Academic	27	30	33	48	56	19	7	11	15	30
Aerospace	20	50	45	60	55	30	15	20	40	25
Auto	10	40	60	80	80	20	40	20	50	50
Energy	20	70	50	65	50	15	45	10	50	20
Enter.	6	67	33	67	50	67	17	17	50	33
Financial	20	65	80	55	35	20	20	5	35	30
Gov. Lab	34	35	32	41	38	29	15	9	24	29
Insurance	5	100	80	60	20	0	20	0	20	0
Robotics	12	33	33	50	50	25	17	8	33	50
Semi-con	8	50	25	62	25	38	0	12	12	0
Tele-com	12	58	58	50	42	17	17	33	33	17
Other	56	48	38	54	36	20	21	9	32	41

Next we looked into how application area varied by industry to see which combinations stand out (Table 13-2). Some specific combinations that we found are that those working in the automotive and robotics industries are interested in multi-objective and numerical optimization problems, while people working in the energy and entertainment industries are interested in multi-objective and classification problems. Finally, those working in insurance, telecommunications and the financial industries are predominantly interested in classification and data mining.

6. CLUSTER SIZE

Computer power can impact the applicability of EC on certain applications. We summarize the computer cluster size used by participants in their jobs as follows: 54.2% uses no cluster (single computer), 15.3% uses 3-10 processors, 24.8% uses 11-100 processors, 4.0% uses 101-1000 processors, and 0.7% uses more than 1000 processors. In general, there are no significant differences in cluster size between academics and non-academics. One exception is in the 11-100 processor range, where 27.1% of academics use a cluster of 11-100 processors while only 20.5% of practitioners in industry do.

When computer cluster sizes were analyzed with problem type, Table 13-3 shows no significant pattern. However, when they are compared with application area, Table 13-4 gives some interesting trends. A couple of trends that can be seen are that the automotive and aerospace industries use large clusters of 101-1000 computers more than other application areas do and that the insurance industry uses only a single computer or two. Two respondents use cluster sizes of more than a thousand processors, of which one is a non-academic consultant

Table 13-3. Percentages of problem types using particular cluster sizes.

Problem type	Total #	Percentage using this cluster size.				
	Jobs	1-2	3-10	11-100	101-1000	1001+
Classification	156	57	16	20	5	0
Data mining	130	53	16	24	5	0
MOO	163	48	19	28	3	0
Num. Opt.	154	47	18	30	3	0
Open Design	88	37	21	34	6	0
Planning	56	46	25	28	0	0
Sat/TSP	41	46	14	34	4	0
Scheduling	88	43	20	32	3	0
Other	143	51	12	26	7	1

Table 13-4. Percentages of job areas using particular cluster sizes.

Job Area	Total #	Percentage using this cluster size.				
	Jobs	1-2	3-10	11-100	101-1000	1001+
Academic	282	55	12	27	3	0
Aerospace	30	41	20	20	17	0
Auto	16	6	26	40	26	0
Energy	33	51	33	15	0	0
Enter.	14	28	50	14	7	0
Financial	25	60	16	24	0	0
Gov. Lab	57	48	12	32	7	0
Insurance	6	100	0	0	0	0
Robotics	36	30	22	41	5	0
Semi-con	11	63	18	18	0	0
Tele-com	18	61	16	22	0	0
Other	82	50	25	20	2	1

in Europe using it for agent-based simulations and the other is an academic in North America using it for classification and information retrieval.

7. EC ACCEPTANCE IN INDUSTRY

Next we examined non-academic jobs to see what trends exist in the distribution and acceptance of EC in industry. Even though there is an exponential growth in the number of yearly EC positions, the ratios between the different levels of distribution and acceptance has remained fairly constant throughout the years. The acceptance rate has averaged: 41.3% well accepted; 19.8% accepted; 36.9% somewhat accepted; and 2.0% rejected. The distribution rate has averaged: 36.4% well distributed; 12.3% distributed; 25.3% somewhat distributed; and 26.0% isolated. That these ratios have remained fairly constant over the years does not mean that EC is not becoming more distributed and accepted in non-academic organizations – in fact, the growth in the number of EC positions implies the opposite. What we cannot determine from our data is whether there is an increase in acceptance and distribution within an organization over time, and this is a question for a future survey.

We also analyzed EC acceptance in non-academic organizations by geographic region. The breakdown of acceptance in Asia, Europe and North America is as follows (well accepted, somewhat accepted, not well accepted, rejected): Asia (53%, 7%, 40%, 0%); Europe (41%, 35%, 20%, 4%); and North America (42%, 21%, 34%, 3%); We do not give a breakdown for the other geographic regions due to insufficient responses.

To increase the acceptance and distribution of EC in industry, it is important to understand the obstacles to its uptake. Based on our responses, we found the obstacles to be: *poorly understood* (39.7%), *too ad hoc* (22.5%), *few successful applications to convince management* (21.2%), *commercial tools were unavailable or ineffective* (20.5%), *Other* (18.5%), *no proof of convergence* (14.6%), and *too hard to apply* (13.9%). In some ways, it is encouraging that the main obstacle is that EC is *poorly understood* because as more universities teach EC techniques, these methods should grow in familiarity and thereby gain wider acceptance in industry. Similarly, with a growth in familiarity of EC, companies may be less inclined to find it "ad hoc". The third main obstacle is *the lack of successful applications*, is being addressed through Real-World Applications tracks at EC conferences and with the Human Competitive Competition held at GECCO since 2004. Finally, *lack of useful commercial tools* suggests a possible market niche for those wanting to achieve commercial success with creating EC software. Among the 27 responses for *Other*, the most common obstacles were: lack of experience/familiarity (9), and too slow or does not scale (4).

8. COMMENTS FOR FUTURE SURVEYS

Having conducted the first survey of practitioners of evolutionary computation we have some thoughts on changes that should be done for future surveys.

First off, to better understand EC education in universities, it would be useful to ask for each degree received the number of courses taken which EC techniques were taught. This would be beneficial for finding out how wide-spread EC techniques are being taught to non-EC specialists and also to find out if EC is being more widely included in the course curriculum. Similarly, it would be useful to query people as to how many EC-specialized conferences they have attended in a given year, or the average number of such conferences they attended a year over the course of each job. This helps us to evaluate if EC-conferences are helpful in disseminating and educating EC technology for practitioners in industry.

Second, in our questions on asking how well accepted/distributed EC is at a particular company, rather than having categories such as “Well accepted” to “Rejected or poorly accepted” for possible answers it would be more useful to ask for a numerical rating from 1 to 5, or 1 to 10 asking for the degree of acceptance. In this case 1 would be “Rejected” and the highest value would be “Well accepted.” Such a numerical system would allow for more fine-grained ranking of acceptance and would allow for numerical processing on how acceptance has changed. Also, it would be useful to ask for the level of EC acceptance at the start of a job and the level of EC acceptance at the end of the job (or its current level of acceptance for jobs in which the respondent is still currently employed at). This would allow for analyzing whether there has been an increase in acceptance of EC at individual companies over time. Another question of use would be to ask for the size of the company or organization which the user is working at. It would be interesting to see if there are trends in the size of organization that uses EC, or in its growth in acceptance.

Finally, in addition to canonical evolutionary algorithms (such as genetic algorithms and evolutionary strategies) in recent years various other biologically-inspired computing algorithms, such as ant colony optimization, artificial immune systems and particle swarm optimization, have been developed. It would be useful to add a question asking respondents about which techniques they have used at each of their jobs so as to track their use and also learn what applications they are being used for.

9. CONCLUSION

Over the years, the use of EC techniques have grown from a few isolated practitioners into a genuine field with a large community. This first survey on EC practitioners has provided us with a preliminary picture of its development in the world. There has been an exponential growth in the number of EC practitioners and EC-specialized graduates, since the first graduates with EC-specialized degrees appearing in the mid 1990’s. After graduation, most survey participants found their jobs through networking or from their supervisors. Encouragingly, along with the growth in EC positions has been a growth in acceptance of EC techniques in industry, with the main obstacle to industry acceptance being that the technique is not well understood. EC has been applied to a wide variety of application areas and different problem domains, among which the most common problem areas are multi-objective optimization, classification, and numerical optimization. Although there are still challenges to the continued transfer of Evolutionary Computation to industry, we hope that the results of this survey will help.

Acknowledgments

We would like to thank Gerardo Valencia for putting together the survey website, and also Erik Goodman, David Goldberg, John Koza and Una-May O'Reilly for their support in conducting this survey.

An abridge version of this article was printed in SIGEVOlution, the newsletter of the ACM Special Interset Group on Genetic and Evolutionary Computation, Volume 2 Issue 1, 2007.

References

R. Atkinson and J. Flint. (2001) Accessing hidden and hard to reach populations: Snowball research strategies. *Social Research Update*, 33, 2001.

K. A. De Jong. (2006) *Evolutionary Computation: A Unified Approach*. MIT Press, 2006.

D. B. Fogel, editor. (1998) *Evolutionary Computation: The Fossil Record*. IEEE Press, Piscataway, NJ, 1998.

W. P. Vogt. (1999) *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Science*. Sage, London, 1999.