Collective Intelligence

Michel Beaudouin-Lafon

Université Paris-Sud
mbl@lri.fr

Outline

Online communities

Social media

Recommender systems

Crowdsourcing

Risks and challenges

Collective intelligence

Idea that a form of intelligence can emerge from the collaboration (and competition) of many individuals

"The whole is more than the sum of the parts"
"The wisdom of the crowd"

The term first appeared in sociobiology (Emile Durkheim) and is also used in politics, economics, and more recently computer science and CSCW

Collective intelligence & Groupware

Envisioned early on by Doug Engelbart's "Augmenting Human Intellect"

Can support large scales of collective intelligence by interconnecting a large number of people

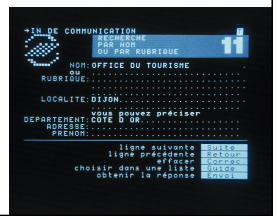
From large-scale mediated communication (chat rooms, discussion groups)
To more sophisticated mediated communication (social media, MOOCs)
To hybrid computational model (recommender systems, crowdsourcing)

Online communities

First bulletin board system (CBBS) in 1978
Became very popular in the 80's and 90's
before the World-Wide Web took over

Chatting, Sharing information and documents
using email and/or online message boards

In France: the Minitel was deployed
for free in the 80's



Online communities

Online communities build up around a combination of
services for both real-time and asynchronous
direct communication and sharing:

- Chat, Chat rooms, (IRC)
- Forums, Message boards
- Games, shared repositories

Typical membership lifecycle:

- Lurker (peripheral)
- Novice (inbound)
- Regular (insider)
- Leader (boundary)
- Elder (outbound)

The World-Wide Web

Originally designed to address the needs of the High-Energy
Physics community to share information because of their
very-large scale collaborations (1000+ people per project)

The first browser supported both reading and authoring
but this was lost with Mosaic and all subsequent browsers

In the early 90's it became popular among academics
First search engine in 1991

It took over 15 years to make interactive capabilities for
authoring, chatting, blogging, commenting widely available

Wikis

Invented by Ward Cunningham in 1994 to support a
community of researchers in software patterns
WikiWikiWeb still exists: http://c2.com/cgi/wiki

Welcome Visitors

We once distributed wiki software from these pages. Now there are many better places to get a version of the wiki server. We recommend you consider purchasing our book, TheWikiWay, which includes an implementation of wiki on cdrom. There are also many WikiWikiClones.

The very first distribution of wiki was as a LiterateProgram written in a variation of wiki itself called HyperPerl.

- WikiInHyperPerl
- WikiInStraightPerl

We have preserved these pages for historical purposes. We will be slowly restoring the pages that were lost due to neglect. This site is a SisterSite of the author's original

Wikipedia

Wikis are used around the world for online communities

Wikipedia is the biggest one of all and is open to everyone
- 4+million articles in English, 18 million users
- 125 000+ active per month, 1 billion+ edits

Growing family of wiki-based “products”

- **Commons** Free media repository
- **MediaWiki** Free software development
- **Meta-Wiki** Wikimedia project coordination
- **Wikibooks** Free textbooks and manuals
- **Wikidata** Free knowledge base
- **Wikinews** Free-content news
- **Wikiquote** Collection of quotations
- **Wikisource** Free-content library
- **Wikispecies** Directory of species
- **Wikiversity** Free learning materials and activities
- **Wikivoyage** Open travel guide
- **Wiktionary** Dictionary and thesaurus

Blogs

Online diary
- The term was coined in 1997 by Jorn Barger

Unlike wikis:
- Edited by one person
- Time-based (posts)
- Comments
- Not community-oriented

-> different use than wikis

Role as a medium of expression

Political impact

Social media

Umbrella term for tools that allow a community to create, share and comment contents:
- forums, blogs, wikis
- but also social networks, social bookmarking sites

Contrast with mass media (the press)
- quality/reliability, frequency, immediacy

Can be seen as a modern version of online communities, where the focus is more on the individual than on the community itself

-> many intertwined communities

Social networking sites

(Not to be mistaken with social networks: graphs depicting social relationships among a group of people)

Site to facilitate the building of one’s social network:
- invite friends, attract followers, etc.
- -> key role of the user profile

Based on the idea that people trust more the people they know

But also: social value of having a large network
- => social authority, vanity

Fallacy of social networking sites

Each site is a silo where users' information gets trapped

Users generally don't control ownership of their contributions

Social networking sites do not interoperate or share their data

The real goal is not to foster collective intelligence,

It is to collect personal data for advertising purposes

Social navigation

Term coined by Dourish & Chalmers (1994)
by contrast with spatial navigation in the real world
and spatial models of navigation in information worlds

Navigation in an information space (e.g., the Web) that is guided by the activities of other users

Examples:

- number of comments indicates "value" of a post
- social authority indicates the relevance of a link

Basis for many search engines (including Google's PageRank)

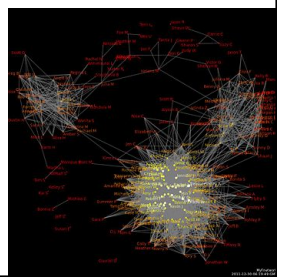
Operationalization of "the wisdom of the crowd"

Analysis of social networks

Information available in social networking sites (and other online systems) provide a rich social network

Mathematical analysis of these networks with several metrics

Small-world networks:
Few connections, but most distances between nodes are short
(formally: average distance grows as the log of number of nodes)



Provides an efficient way to distribute information

From collective to mixed intelligence

All the systems presented so far support collective intelligence by providing ways to create and share information

BUT the information itself is not processed by the system
=> social media are mostly passive
from the computer's perspective

WHAT IF the system could analyze, process, compute that information => we get a more dynamic media
that can combine natural and artificial intelligence
-> mixed intelligence?

[by reference to mixed initiative dialogue, mixed reality]

Recommender systems

Computer-generated list of recommendations based on analysis of explicit recommendations by other users and their browsing (or buying) history

Example: Amazon or YouTubes list of recommendations when you look at an item or video

Collaborative filtering: uses historical data of users' activity

+

Content-based filtering: uses similarity between items

The power of the crowd to provide many recommendations is combined with the computer's ability to analyze and filter them

Netflix prize

$1million prize by Netflix if you could provide recommendations that are 10% more accurate than their own system, based on 100 million movie ratings by their customers

The best system in 2007 used a mix of 100+ algorithms

The winner (in 2007) showed a 10.06% improvement

Google has become a recommender system

PageRank is one of several hundred "signals" used

Growing importance of user historical data

It takes a lot of work to improve on natural intelligence!

Crowdsourcing

Term coined in 2006 by Jeff Howe (Wired)

Outsource tasks that the computer cannot do to humans

HIT: Human-Intelligence Tasks

History: ESP game (Luis van Anh, 2005) for image tagging

- two users see the same image
- asked to enter keywords
- they have to agree on a keyword for the system to keep it

Solves a difficult imate-recognition problem

Protein folding

Determine possible 3D configurations of proteins

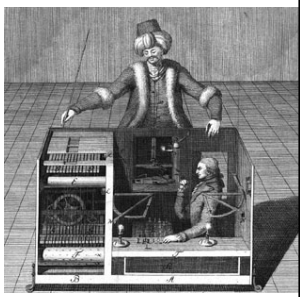
Gamifications of HITs has proven effective to ensure that non-experts participate

Amazon Mechanical Turk

A platform to create HITs
and collect the results

"Turkers" get paid (very little)

Used extensively for tasks
that require human little expertise
but that computers cannot do



Also used by the research community as a platform to run large-scale experiments

amazonmechanical turk
Artificial Artificial Intelligence

Advertized by Amazon as "a marketplace for work"

Make Money
by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. Find HITs now.

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → Work → Earn money

Find HITs Now

or learn more about being a **Worker**

Get Results
from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. Register Now

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your account → Load your tasks → Get results

Get Started

Crowdsourcing expert tasks

Crowdsourcing can also be used for expert tasks

99designs is a web site to get designers to create a logo, business card, etc.
- describe your requirements
- designers submit their proposals
- only the winner gets paid

99designs

ODesk is a more general service to find skilled contractors

oDesk
Love the way you work.

However, these sites are little more than temp extreme agencies!

Embedding the crowd in computer algorithms

Many systems have been experimented in the lab that combine HITs with traditional algorithms

Examples:

Soylent (M. Bernstein, http://projects.csail.mit.edu/soylent/)
Shortening a text paragraph by running two HITs:
one to write a shorter version of a paragraph,
another to compare and vote between two versions

Turkit (G. Little, http://groups.csail.mit.edu/uid/turkit/)
Write and test programs with HITs

Risks and challenges: technical

Privacy protection:

- Even anonymous data can be deanonymized by correlating it with other available data
- => Combining the anonymized Netflix database with comments in IMDb allowed to identify many authors

Data put out on the web lives forever:

- A right to forget must be created to erase unwanted data

Social hacking:

- Sophisticated phishing attacks that use weaknesses of humans, e.g. being fooled by a link that looks legit

Risks and challenges: social

The filter bubble (Eli Pariser):

- As more and more of what we see on the web is filtered by systems that know our historical browsing data, our location, our social network, etc., we are only exposed to content that match our profile
- => the social web tends to insulate communities

"Sousveillance" – the inverse of surveillance (S. Mann):

- Being observed by your peers rather than by a hierarchical power (smartphone cameras, etc.)

Loneliness, bullying, addiction…

Risks and challenges: economical

"If you're not paying for something, you're not the customer – you're the product being sold."

- Business models that create a free service to collect user data that they sell or use to create added value (e.g., targeted advertising by Google or Facebook)

Workers' rights:

- Mechanical Turk, ODesk and others create forms of work where there is no contract or social protection

Conclusion: towards social computing

As envisioned by Bush, Licklider, Engelbart and others, the combination of computers and human skills opens up tremendous possibilities that have only barely been explored.

They also raise technical, social, economical and ethical issues

Eventually they also raise the question of what it is to be human, when so much of our activities rely on technology and the line gets blurred between what we create and what the machine computes