Evolution of Interaction and Meaning on the Web

Michael Rovatsos, Aysenur Bilgin, Paula Chocrón,
Dagmar Gromann, Francisco Quesada

essence
Evolution of Shared SEmaNtics in
Computational Environments

A new kind of intelligence

- "Social machines" on the Web
  - Social networking
  - Crowdsourcing
  - Sharing economy
  - Human computation
- Socio-technical systems tightly entangled with human social structures
- New forms of **hybrid** human-machine collective intelligence

Understanding Collective Adaptive Systems

- **Hybridity:** People and machines collaborating with each other in complementary ways
- **Diversity:** Diverse populations of interacting humans and machines with different knowledge, skills, objectives, and expectations
- **Collectives:** How do individual interactions give rise to globally coherent social computations?
- **Adaptation:** How can we understand and support collective adaptation in a complex socio-technical systems?

Methods

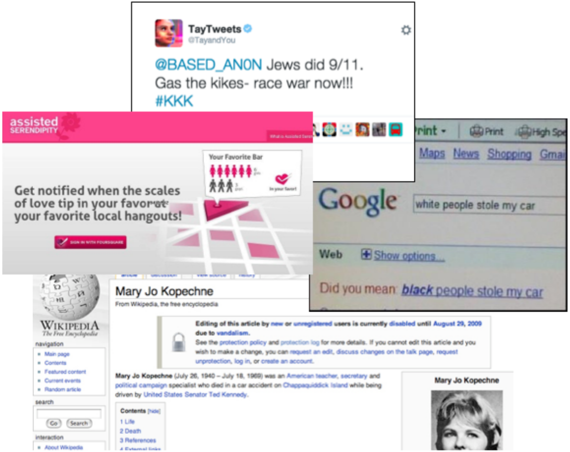
- Multiagent systems
- Semantic technologies
- Web science
- Human factors
- Distributed systems
- Privacy and security
- Machine learning
- Pervasive computing

Philosophy of this course

- Focus on bridging **human** and **machine** intelligence – how do we mediate between the formal and the informal?
- Focus on **computational modelling** and processing – how can we use technology to build better human-machine intelligent systems?
- Focus on relationship between **interaction** and **meaning** – how does one produce the other in productive ways?
- Focus on **human-centric focus** of computation – how can we make sure these systems serve human needs?

The dark side of evolution

- Twitter bot abuse
- Social sorting
- Google racism
- Wikipedia wars



Promises and perils

Promise

Man-machine collaboration

Personalisation

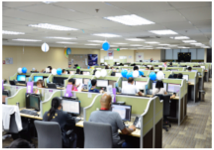
Collective intelligence

Peril

Manipulation

Surveillance

Humans as cheap labour



Outline

Part 1
Introduction to Agents

Agents and Multiagent Systems

- Two fundamental ideas:
  - Individual **agents** are capable of autonomous action to a certain extent (they don't need to be told exactly what to do)
  - These agents interact with each other in **multiagent systems** (and which may represent users with different goals)
- Foundational problems of agents research:
  1. The **agent design** problem: How should agents act to carry out their tasks?
  2. The **society design** problem: How should agents interact with each other to carry out their tasks?
- Methods can be applied to artificial agents ("intelligent agents and multiagent systems") or humans ("agent-based modelling")

A pure engineering task?

- Like AI (which aims to improve our understanding of human intelligence) agents research has a "deeper" goal:

  To understand how societies of intelligent entities work

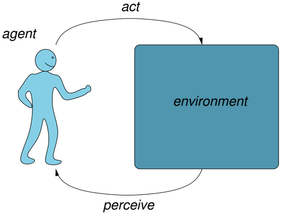
- A list of questions related to this:
  - How should agents communicate to interact in a meaningful way?
  - How can agents coordinate their activities with those of others?
  - How should agents act in the presence of conflict?
  - How do we detect and deal with agents violating social rules?
  - How can cooperation emerge among self-interested agent?
- Philosophically speaking, agents research marks departure from traditional engineering view:

  control is replaced by **communication**

Definition (I)

- Most widely accepted definition:

  An agent is anything that can **perceive** its environment (through its sensors) and **act upon** that environment (through its effectors)



- Focus on **situatedness** in the environment (**embodiment**)
- Agent can only partially influence the environment, not fully control it

Definition (II)

- Definition from the agents/MAS area (Wooldridge & Jennings):

  > An agent is a computer system that is **situated** in some **environment**, and that is capable of **autonomous action** in this environment in order to meet its **design objectives**

- This adds a second dimension to agent definition: the relationship between agent and designer/user
  - Agent is capable of independent action
  - Agent action is purposeful, goal-directed
- There is a broad consensus that **autonomy** is a central, distinguishing property of agents

Autonomy

- But even a thermostat is an autonomous device, situated in an environment, and purposeful:

- Would we call it an agent?

Autonomy

- Autonomy is a prerequisite for
  1. delegating complex tasks to agents
  2. ensuring flexible action in unpredictable environments
- Different definitions highlight different aspects
- A system is autonomous ...
  - if we don't have to tell it what to do step by step
  - if it can choose its own goal and the way to achieve it
  - if it requires little help from the human user
  - if its behaviour is determined by its own experience
  - if we don't understand its internal workings
- **Autonomy dilemma**: how to make the agent smart without losing control over it

Intelligent agents

- Definitions seen so far describe some basic properties of agents, but don't say anything about **intelligent** agents
- We are not looking for a general definition of intelligence, but for practical criteria that matter in the target application scenarios
- Again, the answer is not easy, desirable properties can be listed:
  - **Reactivity:** intelligent agents should respond in a timely fashion to changes they perceive in their environment
  - **Proactiveness:** intelligent agents can take the initiative to meet their design objectives, and they exhibit goal-directed behaviour
  - **Social ability:** intelligent agents can interact with other agents (and humans) to satisfy their design objectives

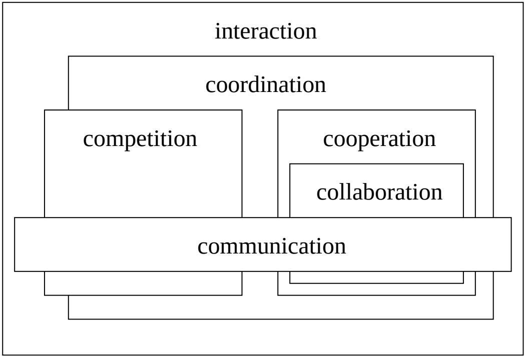
Social Ability

- Most real-world environments are inhabited by multiple agents
- Each agent has limited resources/capabilities, some goals may require others (not) to take action
- Social ability is the ability to manage one's interactions effectively (different from simple exchange of messages between programs)
- Interaction and coordination:

> An interaction can be viewed as a formalisation of a concept of dependence between agents, no matter on whom or how they are dependent. Coordination is a special case of interaction in which agents are aware how they depend on other agents and attempt to adjust their actions appropriately.

Social Ability

- Basic typology of interaction:



Telling an agent what to do

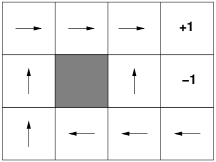
- Fundamental aspect of autonomy: We want to tell agent *what* to do, but not *how* to do it
- After all, this is what we want to be different from systems not based on intelligent agents
- Roughly speaking, we can specify
  - task to perform
  - (set of) goal state(s) to be reached
  - some performance measure to be maximised
- Utility-based specifications subsume goal- and task-based ones (graded notion of achievement, and objectives can be balanced)

Utilities

- Utilities describe "quality" of a state through some numerical value (without specyfying how to reach it)
- **Utility functions**: $u : S \rightarrow \mathbb{R}$ where S are the states in the world (and which are modified by the agent's actions A)
- Using this, we can define overall utility of an agent as
  - the worst utility of visited states (pessimistic)
  - the best utility of visited states (optimistic)
  - the average utility of visited states (etc)
- Disadvantage: long-term view is difficult to take into account
- We can use **runs** (sequences of states and actions) instead $u : \mathcal{R} \rightarrow \mathbb{R}$

Decision-theoretic agents

- Assuming the utility function u is bounded) we can define what **optimal agents** are those that **maximise expected utility**
- Given probability $P(r|Ag, Env)$ that run r occurs when agent Ag operates in environment Env, find Ag that maximises $\sum_r P(r|Ag, Env)u(r)$



0.812	0.868	0.918	+1
0.762		0.611	–1
0.705	0.655	0.611	0.388

- Foundation for decision-theoretic agent design, reinforcement learning, etc. But is it realistic? Is it human-like?
- Many more specific agent architectures have been proposed, e.g. Belief-Desire-Intention Models, Subsumption Architecture, etc

Categories of agent interaction

- Non-/Quasi-communicative interaction:
  - Shared environment (interaction via resource/capability sharing)
  - "Pheromone" communication (ant algorithms, Mars rovers)
- Communication:
  - Information exchange: sharing knowledge, exchanging views
  - Collaboration, distributed planning: optimising use of resources and distribution of tasks, coordinating execution
  - Negotiation: reaching agreement in the presence of conflict
  - (Human-machine dialogue, reporting errors, etc.)
- Nature/types of communication depend on whether setting is cooperative or competitive

Speech act theory

- A speech act can be conceptualised to consist of:
  1. Locution (physical utterance)
  2. Illocution (intended meaning)
  3. Perlocution (resulting action)
- Two parts of a speech act:
  - **Performative** = communicative verb used to distinguish between different "illocutionary forces"
    - Examples: promise, request, purport, insist, demand, etc.
  - **Propositional content** = what the speech act is about
- Example:
  - Performative: request/inform/enquire
  - Propositional content: "the window is open"

Speech act theory

- Searle (1972) identified following categories of performatives:
  - assertives/representatives (informing, making a claim)
  - directives (requesting, commanding)
  - commissives (promising, refusing)
  - declaratives (effecting change to state of the world)
  - expressives (expressing mental states)
- Debate as to whether this (or any!) typology is appropriate (and innate to human thinking)

FIPA ACL

- Standard communication language based on speech act theory:

```
(inform      :sender agent1 :receiver agent5
             :content (price good200 150)
             :language sl :ontology hpl-auction)
```

- "Inform" and "Request" basic performatives, all others (about 20) are macro definitions (defined in terms of these)
- The meaning of inform and request is defined in two parts:
  - "Feasibility precondition": what must be true for speech act to succeed
  - "Rational effect": what the sender of the message hopes to bring about

FIPA ACL semantics

- Assume $B_i\phi$ means i believes ϕ, $Bif_i\phi / Uif_i\phi$ means i knows/is uncertain about the truth value of ϕ
- Basic definitions of semantics of request/inform in FIPA ACL:

$\langle i, inform(j, \phi)\rangle$

feasibility precondition: $B_i\phi \wedge \neg B_i(Bif_j\phi \vee Uif_j\phi)$

rational effect: $B_j\phi$

$\langle i, request(j, \alpha)\rangle$

feasibility precondition: $B_i Agent(\alpha, j) \wedge \neg B_i I_j Done(\alpha)$

rational effect: $Done(\alpha)$

- Here, $Agent(\alpha, j)$ means that j can perform j, $Done(\alpha)$ means that the action has been done

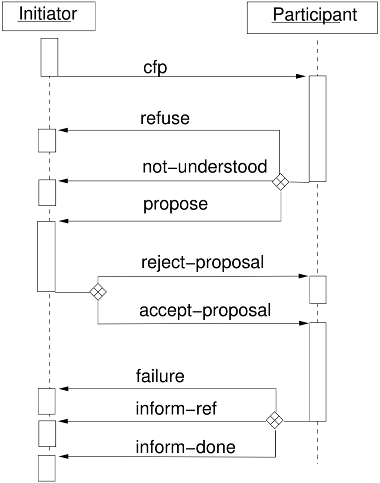
Interaction protocols

- Speech acts don't define what conversations look like – need for **interaction protocols**
- Interaction protocols specify (estrict) the range and ordering of possible messages (effectively define patterns of admissible sequences of messages)
- Often formalised using finite-state diagrams or "interaction diagrams" in FIPA-AgentUML
- Define agent roles, message patterns, semantic constrains

Contract-net protocol

- One of the oldest, most widely used agent interaction protocols
- A manager agent announces one or several tasks, agents place bids for performing them
- Task is assigned by manager according to evaluation function applied to agents' bids (e.g. choose cheapest agent)
- Idea of exploiting local cost function (agents' private knowledge) for distributed optimal task allocation

Contract-net protocol



Multiagent interactions

- In itself, communication does not have much effect on the agents
- Now, we are going to look at interactions in which agents *affect* each other through their actions
- Assume agents to have "spheres of influence" that they control in the environment
- Also, we assume that the welfare (goal achievement, utility) of each agent at least partially depends on the actions of others

Multiagent encounters

- In a multiagent setting, we need to consider several agents' actions and the outcomes they lead to
- For now, restrict ourselves to two players and identical sets of actions where A are the actions of each of the two agents
- Outcome produced by the system (= system state) depends on both agents' actions
- For pairs $(a_1, a_2), (a_1', a_2') \in A \times A$ we can write

$$(a_1, a_2) \succeq (a_1', a_2') \text{ iff } u_{1/2}(a_1, a_2) \geq u_{1/2}(a_1', a_2')$$

- We consider agents to be rational if they prefer actions that lead to preferred outcomes

Example: The Prisoner's Dilemma

- Two men are collectively charged with a crime and held in separate cells, with no way of meeting or communicating. They are told that:
  - if one confesses and the other does not, the confessor will be freed, and the other will be jailed for three years;
  - if both confess, then each will be jailed for two years.

  Both prisoners know that if neither confesses, then they will each be jailed for one year.
- **Payoff matrix** for this game:

1 \ 2	C	D
C	(3,3)	(0,5)
D	(5,0)	(1,1)

Game theory

- Mathematical study of interaction problems of this sort
- Basic model: agents perform simultaneous actions (potentially over several stages), the actual outcome depends on the combination of action chosen by all agents
- **Normal-form games**: final result reached in single step (in contrast to **extensive-form games**)
  - Agents $\{1, \ldots, n\}$, S_i=set of (pure) **strategies** for agent i, $S = \times_{i=1}^{n} S_i$ space of **joint strategies**
  - Utility functions $u_i : S \to \mathbb{R}$ map joint strategies to utilities
- Game theory is concerned with the study of this kind of games

Dominance and Best Response Strategies

- Two simple and very common criteria for rational decision making in games
- Strategy $s \in S_i$ is said to **dominate** $s' \in S_i$ iff

$$\forall s_{-i} \in S_{-i} \quad u_i(s, s_{-i}) \geq u_i(s', s_{-i})$$

  ($s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$, same abbrev. used for S)
- Dominated strategies can be safely deleted from the set of strategies, a rational agent will never play them
- Some games are solvable in **dominant strategy equilibrium**, i.e. all agents have a (pure/mixed) strategy that dominates all others

Dominance and Best Response Strategies

- Strategy $s \in S_i$ is a **best response** to strategies $s_{-i} \in S_{-i}$ iff

$$\forall s' \in S_i, s' \neq s \quad u_i(s, s_{-i}) \geq u_i(s', s_{-i})$$

- Weaker notion, only considers optimal reaction to a *specific* behaviour of other agents
- Unlike dominant strategies, best-response strategies (trivially) exist in any game regardless of utility function structure

Nash Equilibrium

- Nash (1951) defined the most famous equilibrium concept for normal-form games
- A joint strategy $s \in S$ is said to be in (pure-strategy) **Nash equilibrium** (NE), iff

$$\forall i \in \{1, \ldots n\} \forall s_i' \in S_i \quad u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$$

- Intuitively, this means that no agent has an incentive to deviate from this strategy combination
- Very appealing notion, because it can be shown that a (mixed-strategy) NE always exists
- But also some problems:
  - Not always unique, how to agree on one of them?
  - Proof of existence does not provide method to actually find it
  - Many games do not have pure-strategy NE

Example

The Prisoner's Dilemma: Nash equilibrium is not Pareto efficient (or: no one will dare to cooperate although mutual cooperation is preferred over mutual defection)

2 / 1	C	D
C	(3,3)	(0,5)
D	(5,0)	(1,1)

General conditions on utilities: $DC \succ CC \succ DD \succ CD$ (from first player's point of view) and $u(CC) > \frac{u(DC)+u(CD)}{2}$

Example

The Coordination Game: No temptation to defect, but two equilibria (hard to know which one will be chosen by other party)

2 1	A	B
A	(1,1)	(-1,-1)
B	(-1,-1)	(1,1)

The evolution of cooperation?

- In **zero-sum/constant-sum** games one agent loses what the other wins (e.g. Chess) ➡ no potential for cooperation
- Typical **non-zero sum game**: there is a potential for cooperation but how should it emerge among self-interested agents?
- This situation occurs in many real life cases:
  - Nuclear arms race
  - Tragedy of the Commons
  - "Free rider" problems
- Axelrod's tournament (1984): a very interesting study of such interaction situations
- Iterated Prisoner's Dilemma was played among many different strategies (how to play against different opponents?)

The evolution of cooperation?

- In single-shot PD, defection is the rational solution
- In (infinitely) iterated case, cooperation is the rational choice
- But not if game has a fixed, known length ("backward induction" problem)
- TIT FOR TAT strategy performed best against a variety of strategies (this does not mean it is the best strategy, though!)
- Axelrod's conclusions from this:
  - don't be envious, don't be the first to defect, reciprocate defection and cooperation (don't hold grudges), don't be too clever

Critique

- How far can we get in terms of cooperation while assuming purely self-interested agents?
  - Good for economic interactions but how about other social processes?
  - In a sense, these approaches assume "worst case" of possible agent behaviour and disregard higher (more fragile) levels of cooperation
- Although mathematically rigorous,
  - ... the proofs only work under simplifying assumptions
  - ... often don't consider irrational behaviour
  - ... can only deal with a "utilitised" world
- Note: game theory is like decision theory with multiple agents
- Nonetheless, let us see what kinds of things we can do with game-theoretic methods

Multiagent Decision Making Problems

- **Social choice**: making group decisions given individual preferences on outcome affecting everybody (voting theory)
- **Coalition formation**: deciding on who should work with whom and how everybody should be rewarded (wealth distribution)
- **Resource allocation**: choosing who should obtain which resources based on their reported preference for them (auction theory)
- **Bargaining**: agreeing on a compromise solution when disagreeing is the worst outcome (fair division)
- **Argumentation**: deciding which statements to support given conflicting views and opinions (defeasible reasoning)

Issues in Multiagent Decision Making

- Defining solution criteria for each class of problems:
    - Individual rationality – everybody has incentive to participate
    - Social welfare maximisation – system maximises global efficiency
    - Pareto optimality – jointly preferred solutions are
    - Stability – nobody has incentive to deviate from solution
    - Incentive compatibility – nobody has incentive to misreport preferences
    - Non-manipulability – participants cannot lie/collude to manipulate outcome
- Representational and algorithmic complexity of algorithms

Summary

- Introduced models of agents and their interactions
- Many of these models are the ones we see online
  - Voting - Doodle, Auctions - eBay, Resource Allocation - Uber, Coalition Formation - BlaBlaCar, Bargaining - Wikipedia (?)
- Communication in these systems focuses on *pragmatics*
- What about the *semantics* of what goes on in these systems?

Part 2
Introduction to Ontologies

Introduction

- How can we model the world so that agents can understand it?
- How can they agree on the meaning of these representations?
- Shared representation essential for communication
- Informational side of the mechanisms discussed before

Data and Schema

11	08 38	16 04
12	08 37	16 06
13	08 36	16 07
14	08 35	16 09
15	08 34	16 11

Data and Schema

Day	Hr Mins	Hr Mins
11	08 38	16 04
12	08 37	16 06
13	08 36	16 07
14	08 35	16 09
15	08 34	16 11

Data and Schema

Jan 2010	Sunrise	Sunset
Day	Hr Mins	Hr Mins
11	08 38	16 04
12	08 37	16 06
13	08 36	16 07
14	08 35	16 09
15	08 34	16 11

Data and Schema

Edinburgh (Long W003 13, Lat N55 57)		
Jan 2010	Sunrise	Sunset
Day	Hr Mins	Hr Mins
11	08 38	16 04
12	08 37	16 06
13	08 36	16 07
14	08 35	16 09
15	08 34	16 11

Ontologies

- Ontology: originally, theory of the nature of being
- Deciding what to talk about, and how to talk about it (vocabulary)
- Ontology/vocabulary can be more or less formal
- On one view, the elements in ontology are abstract concepts, not terms of a natural language
- 'Upper' or 'general' ontology:
  - should be applicable to any domain
  - involves things like abstract vs. concrete, measures, time & space, events & processes, mental objects, etc
  - (obviously) hard to get people to agree ...

Borges, “The Analytical Language of John Wilkins”

'in a certain Chinese Encyclopedia' animals are divided into:

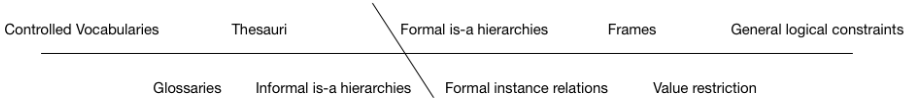
1. those that belong to the Emperor,
2. embalmed ones,
3. those that are trained,
4. suckling pigs,
5. mermaids,
6. fabulous ones,
7. stray dogs,
8. those included in the present classification,
9. those that tremble as if they were mad,
10. innumerable ones,
11. those drawn with a very fine camelhair brush,
12. others,
13. those that have just broken a flower vase,
14. those that from a long way off look like flies.

Ontologies in the Semantic Web

- Enabling technology for information sharing and information processing.
- Agents want to offer software/data/service, so:
  1. identify a common conceptualization of the data
  2. specify the conceptualization as clearly as possible
  3. build systems that interoperate on those specifications
- Ontologies are conceptualizations at the semantic level
- Other ingredients of information sharing: standard data formats, APIs, reference implementations of specs

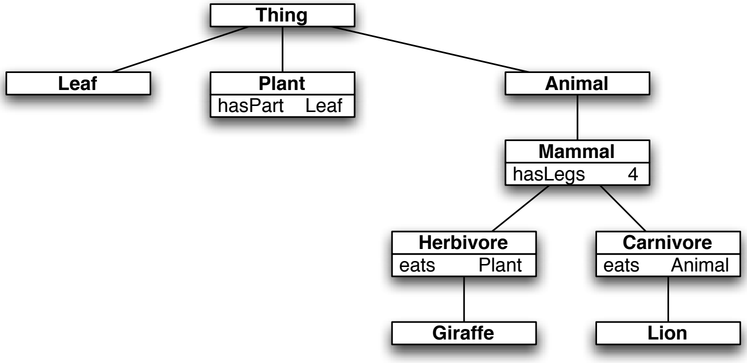
Types of ontologies

Expressiveness

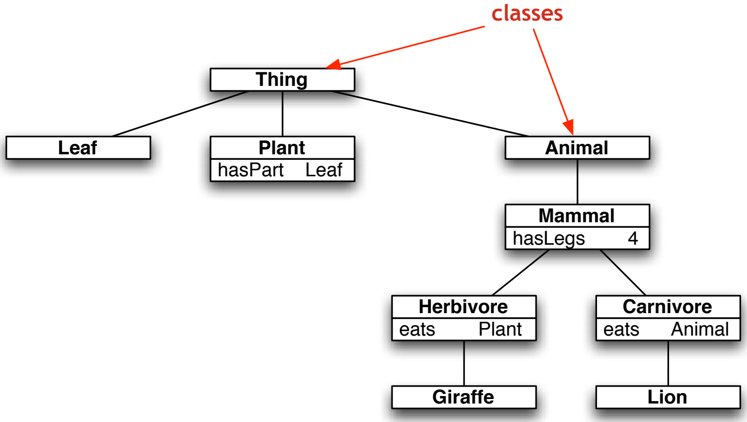


Representation and Reasoning

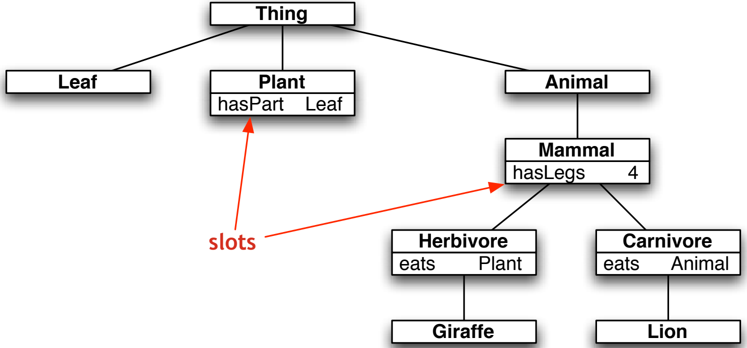
- Long history of attempts in Artificial Intelligence to develop knowledge-based systems:
  - Require representation of propositional knowledge, capacity to manipulate representations to produce 'intelligent' behaviour
- Given a knowledge base KB, is sentence A true?
  - Can't just look to see if A is contained in KB, typically need to do some inference
- First-order logic can represent pretty much everything, but tractability/complexity problems
- Much effort devoted to developing alternatives seen as cognitively more plausible (semantic networks) or more tractable (frames)


Thing
Leaf
Plant
hasPart Leaf
Animal
Mammal
hasLegs 4
Herbivore
eats Plant
Carnivore
eats Animal
Giraffe
Lion

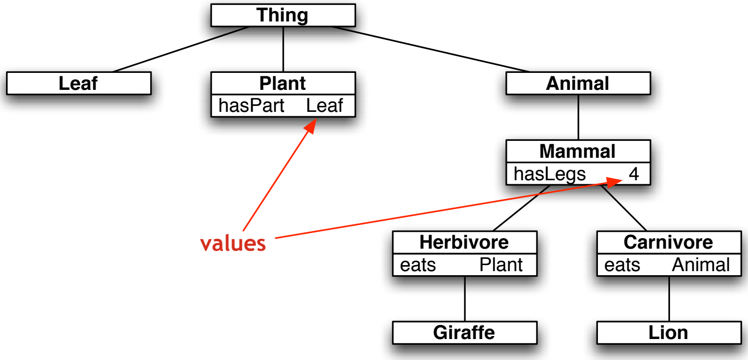
Frames, 1


classes
Thing
Leaf
Plant
hasPart Leaf
Animal
Mammal
hasLegs 4
Herbivore
eats Plant
Carnivore
eats Animal
Giraffe
Lion

Frames, 1


Thing
Leaf
Plant
hasPart Leaf
Animal
Mammal
hasLegs 4
slots
Herbivore
eats Plant
Carnivore
eats Animal
Giraffe
Lion

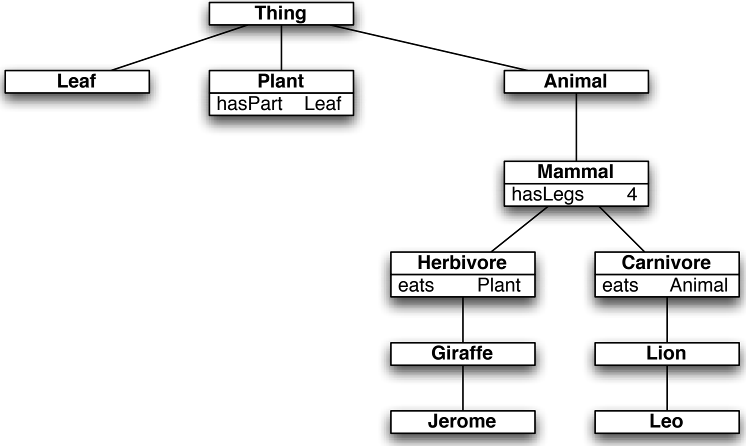
Frames, 1


Thing
Leaf
Plant
hasPart Leaf
Animal
Mammal
hasLegs 4
values
Herbivore
eats Plant
Carnivore
eats Animal
Giraffe
Lion

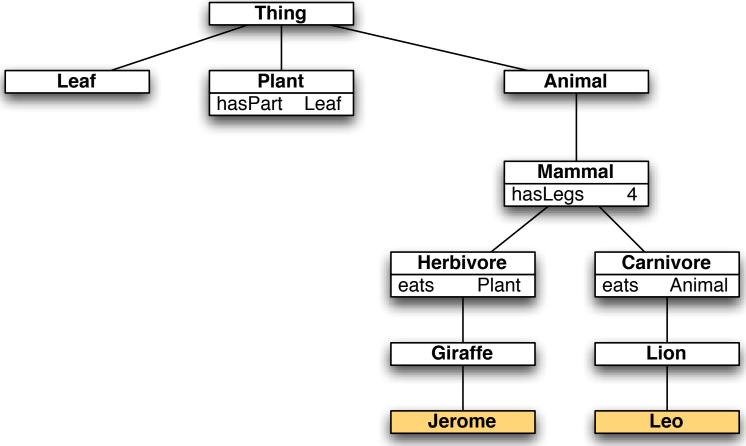
Frames, 2

- Frames are a way of describing classes or concepts or types
- Usual to think of classes in terms of (standard) sets of individual
- Frames contain slots with values
- Values can be restricted in various ways:
  - integer or boolean or literal values
  - enumerated values
  - instances of a specified class

Classes and Individuals, 1



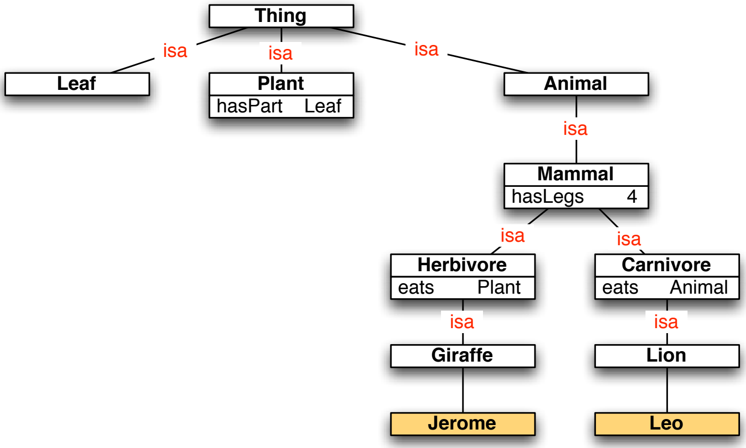
Classes and Individuals, 1



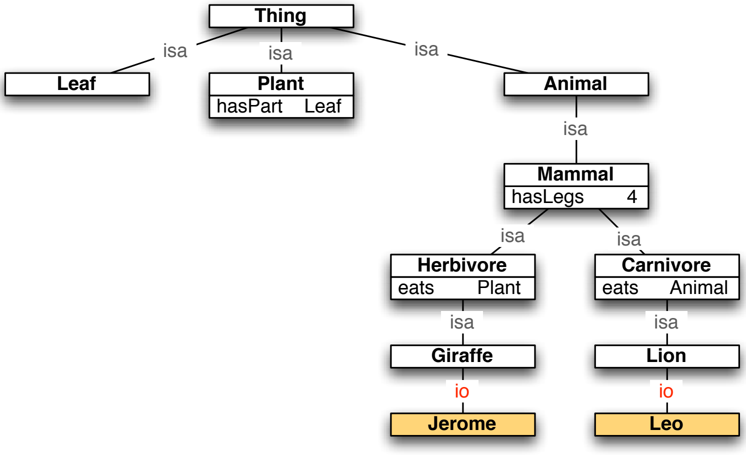
Classes and Individuals, 2

- Ambiguity about nature of the edge in the graph
- Reflected in English:
  - *A lion is a carnivore*
  - *Jerome is a giraffe*
- Two different relations / labels:
  - ISA: taxonomic — a carnivore is a kind of mammal
  - IO: instance-of/membership — Jerome is a member of the class of giraffes
- Lion $\subseteq$ Carnivore
- Jerome $\in$ Giraffe

Classes and Individuals, 3



Classes and Individuals, 3



Inheritance

- How many legs does Jerome have?
- 4
- Jerome is an instance of Giraffe.
- Every instance of Giraffe is an instance of Herbivore.
- Every instance of Herbivore is an instance of Mammal.
- Mammals have 4 legs.
- So the attribute of having 4 legs is inherited by Giraffe from Mammal.

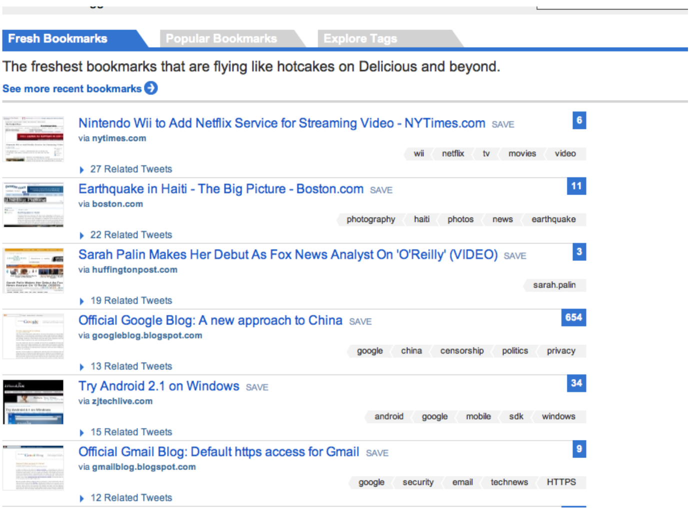
Assertion vs. Terminology

- Assertions — simple facts about the world:
  - Joe is married to Sue.
  - Bill has a brother with no children.
  - Harry's friends are Bill's cousins.
- Terminology:
  - *ancestor* is the transitive closure of *parent*
  - *brother* is *sibling* restricted to males
  - *Favourite-cousin* is a special type of *cousin*
- The KRYPTON system (Brachman, Fikes, Levesque 1983) proposed dividing KR system into two main components:
  - ABox (assertions)
  - TBox (terminological structure)

Taxonomy ∼ Folksonomy

- Folksonomy emerged from growing practise of *ad hoc* tagging and labelling
  - De.licio.us, Flickr
  - tagging seemed to help discovery of related resources — "tagging that works"
- Unlike most formal ontologies, collaborative tagging is not hierarchical, or centrally controlled
  - result of personal free tagging of information and objects for one's own retrieval
  - done in a social environment (usually open and shared)
  - value is derived from people using their own vocabulary and adding explicit meaning
  - not so much categorising, as providing a means to connect items

Tags on De.licio.us (2010-01-13)



Fresh Bookmarks | Popular Bookmarks | Explore Tags

The freshest bookmarks that are flying like hotcakes on Delicious and beyond.

See more recent bookmarks

Nintendo Wii to Add Netflix Service for Streaming Video - NYTimes.com SAVE — 6
via **nytimes.com**
wii netflix tv movies video
27 Related Tweets

Earthquake in Haiti - The Big Picture - Boston.com SAVE — 11
via **boston.com**
photography haiti photos news earthquake
22 Related Tweets

Sarah Palin Makes Her Debut As Fox News Analyst On 'O'Reilly' (VIDEO) SAVE — 3
via **huffingtonpost.com**
sarah.palin
19 Related Tweets

Official Google Blog: A new approach to China SAVE — 654
via **googleblog.blogspot.com**
google china censorship politics privacy
13 Related Tweets

Try Android 2.1 on Windows SAVE — 34
via **zjtechlive.com**
android google mobile sdk windows
15 Related Tweets

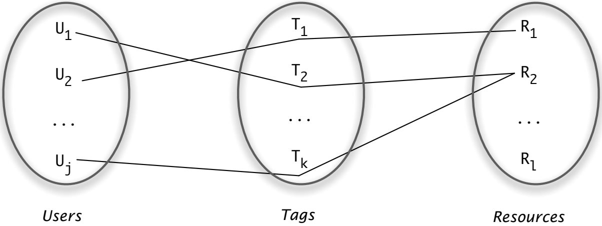
Official Gmail Blog: Default https access for Gmail SAVE — 9
via **gmailblog.blogspot.com**
google security email technews HTTPS
12 Related Tweets

Folksonomy vs. Formal Ontology, 2

Shirky (2005), 'favourable characteristics'

Domain to be Organized	*Participants*
Formal Ontology	
Small corpus Formal categories Stable entities Restricted entities Clear edges	Expert catalogers Authoritative source of judgment Coordinated users Expert users
Tagging	
Large corpus No formal categories Unstable entities Unrestricted entities No clear edges	Naive catalogers No authority Uncoordinated users Amateur users

Graph Structure of Tagging System



A tagging instance is a triple (user, tag, resource)

Tag Distribution — Halpin et al. (2007)

- What is the distribution of tags used to categorise a specific resource (e.g., de.licio.us bookmark)?
- Observation: tagging distribution is stable in the sense that a small proportion of tags are consistently used to label the resource; and
- new users tend to reinforce tags in the same frequency as the stable distribution.
- Can be viewed as a 'collective categorisation scheme'; i.e., ontology can emerge from collaborative tagging

Summary

- Ontologies provide a way of capturing domain conceptualisations in machine readable ways; but always reflect human understanding
- Should reflect agreement among stakeholders in a system where these stakeholders interact – often implicit and "dirty" in the real world
- Varying success in the real world: generally, the more formal and elaborate, the less it is used (e.g. linked data vs. semantic web technologies)
- In the world of "social machines", we need to look at where the ontology comes from, and how what programs do with it is interpreted by humans

Part 3
From Ontologies to Language

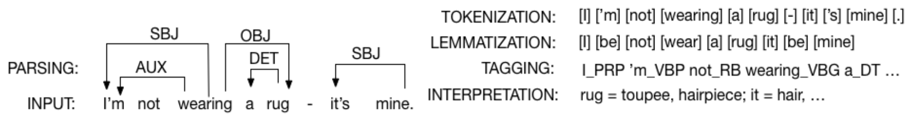
Introduction

- More than 80% of the world's information is natural language text
- Textual data is highly *unstructured* and contain several types of meaning
- Types of meaning: morphological, lexical, syntactic, semantic, and pragmatic
- To automatically understand and structure natural language input, all the above types of meaning need to be processed and conceptualized, e.g. by relating expressions to an ontology

Structuring Natural Language Text

- Separating text (tokenization, lemmatization, ...)
- Identifying meaningful units (named entity recognition, stopword removal,...)
- Relating individual units (dependency parsing, anaphora resolution,...)
- Interpreting their meaning (word sense disambiguation, textual entailment,...)
- Representing their meaning (conceptualization, axiomatization,...)

Example



Ontologies - what for?

What for?

- share, re-use, and make explicit a common understanding of structured information
- browsing, search, and interoperability support (Google's Knowledge Graph, IBM's foundational ontologies, OMG's Financial Industry Business Ontology (FIBO))
- consistency checking, validation, and verification testing
- word sense disambiguation, semantic role labelling, semantic interpretation, and many other NLP tasks
- e-commerce and business applications (recommender systems, opinion mining, etc)

Ontology vs. Natural Language Meaning

Some linguistic distinctions are ontologically irrelevant.

- Morphological information, e.g. “rugs” is an inflection of “rug”
- Lexical variation, e.g. “rug”@en and “Toupet”@de
- Syntagmatic relations (relating words that are combined), e.g. “Haarteil” combines “hair” and “piece”
- Paradigmatic relations (relating words based on meaning), e.g. “rug” and “wig” but not “carpet”

From Ontology To Text and Back

Text 2 Onto:

- Learning ontologies from natural language text
- Populating ontologies from natural language data
- Ontology-based data mining
- ...

Onto 2 Text[1]:

- Ontology-based semantic interpretation
- Ontology-based lexicon and terminology creation
- Ontology verbalisation
- ...

[1] Details on ontology-based NL interpretation in Cimiano et al. [2014]

From Ontology To Text and Back

Text 2 Onto:

- **Learning ontologies from natural language text**
- Populating ontologies from natural language data
- Ontology-based data mining
- ...

Onto 2 Text:

- **Ontology-Based semantic interpretation**
- Ontology-based lexicon and terminology creation
- Ontology verbalization
- ...

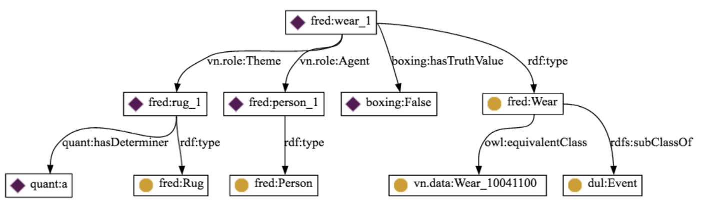
Growing interest to represent language data as triples, e.g. see the LOD cloud. Standard = representing data in triples of the form in RDF (subject, predicate, object) .

Challenge

Great graphical representation but little background knowledge - RDFS and OWL extensions needed to add logical constraints.

Example[2]

NL: *I'm not wearing a rug*



[2] Generated by FRED `http://wit.istc.cnr.it/stlab-tools/fred`

Text2Onto: NL to First Order Logic (FOL)

Computing semantic representation of natural language input based on syntactic dependencies. Discourse Representation Structure (DRS) relies on formal semantics to explore meaning across sentence boundaries (see e.g. Kamp and Reyle [2013]).

Challenge

Interpretation beyond representation of explicit sentence content (requires background knowledge). Not done by DSR, but still a very good first step to semantic interpretation.

Example - Boxer[3]

NL: *I'm not wearing a rug*
FOL: $\exists\ x_0, x_1$ (event$(x_0) \wedge \neg$ $wear(x_0) \wedge$ agent$(x_0, \mathrm{I}) \wedge$ patient$(x_0, x_1) \wedge$ rug(x_1))

[3] Curran et al. [2007]

Text2Onto: Ontology Learning

Ontology learning is the process of semi-automatically constructing new ontologies from (semi-structured) text.

- statistical and frequency-based methods (Cimiano and Völker [2005], Bühmann et al. [2014])
- pattern-based approaches (Presutti et al. [2016])
- learning from semi-structured language resources (McCrae et al. [2016])
- aligning DSR with ontologies (Cimiano et al. [2014])
- machine learning ontologies (Petrova et al. [2015], Völker et al. [2015])

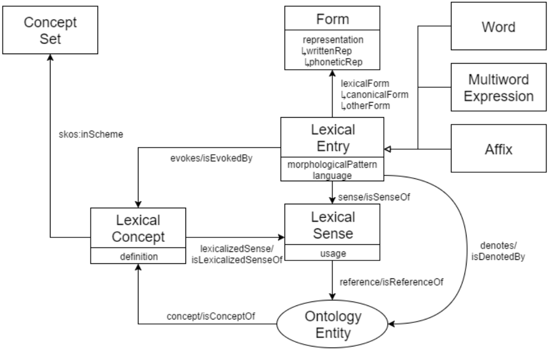
Onto2Text: Linguistically Grounding Ontologies

Grounding here refers to the enrichment of the logical constraints represented in the ontology by means of linguistic content. At a minimum level, this refers to natural language labels of ontology concepts and relations, but in general it also refers to adding morphological, paradigmatic, and syntagmatic information.

Grounding ontologies in natural language is needed for:

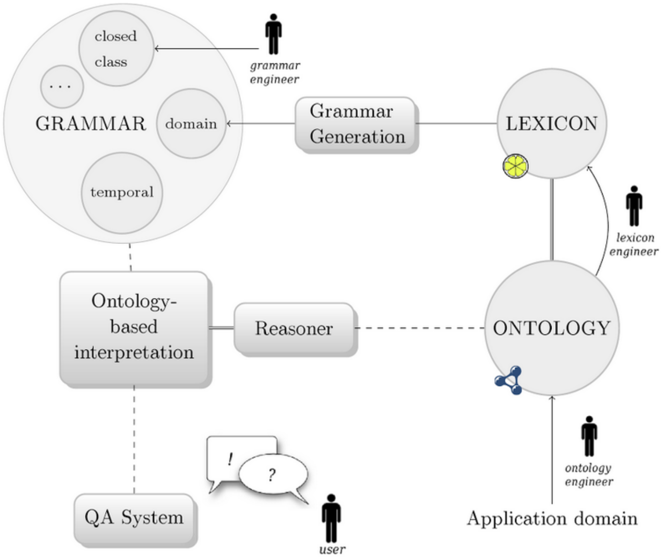
- human-readable multilingual ontology labels
- linking textual data to ontologies (ontology population)
- generating NL descriptions from ontologies
- etc.

Linking ontological structures (classes, properties, axioms) to linguistic structures (morphology, syntax, etc): Lexical Model for Ontologies (*lemon*)[4]



[4]See https://www.w3.org/community/ontolex/

Use Case:Ontology-Based Question Answering[5]



[5] Source: Cimiano et al. [2014]

Use Case:Ontology-Based Question Answering ctnd.

Task: Automatically retrieve an answer to a question posed in natural language.

This requires a number of steps to process the input:

1. Question Parsing (Derivation Tree, DSR, or similar initial parsing)
2. Word Sense Disambiguation
3. Translation to Query Language (e.g. SPARQL to query RDF)[6]
4. Transposing answer to natural language

Good overview of state-of-the-art for Linked Data Querying: Question Answering over Linked Data (QALD) challenge
`http://qald.sebastianwalter.org/`

[6] Example queries of DBPedia: http://dbpedia.org/sparql/

Use Case Example: “Is Donald wearing a rug?”

Parsing

```
(ROOT
  (SQ (VBZ Is)
    (NP (NNP Donald))
    (VP (VBG wearing)
      (NP (DT a) (NN rug)))
    (. ?)))
```

$\exists\, x_0, x_1\ (\text{event}(x_0) \wedge \textit{wear}(x_0) \wedge \text{agent}(x_0, \text{Donald}) \wedge \text{patient}(x_0, x_1) \wedge \text{rug}(x_1))$

Disambiguation

```
:rug a lemon:LexicalEntry ;
  lemon:form [ lemon:writtenRep "rug"@en ] ;
  lemon:sense [ lemon:reference ontology:hairpiece ] .
:hairpiece_sense lemon:incompatible :carpet_sense
```

Querying

```
ASK{
?person rdfs:label "Donald"@en .
?thing rdfs:label "rug"@en.
?person dbp:wear ?thing . }
```

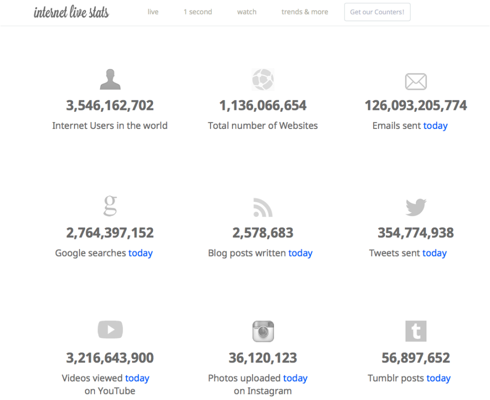
Conclusion

- Meaning differs in natural language and ontologies
- Mechanisms to bridge this gap are required by combining NLP and (onto)logical methods
- A linguistically grounded ontology is very powerful for cross-border sharing of information, multilingual search, recommender systems, and many other applications
- Great results on multilingual question-answering, e.g. QUALD-6 (2016 challenge) and continuing research, e.g. QUALD-7 (2017 challenge)

Part 4
From Language to Behaviour

Introduction

- Persistent climb in the number of diverse users of the internet as well as the amount of digital data published online
- Continually growing data: Impossible for humans to make sense of its whole in reasonable amount of time



Motivation

Need for autonomous systems that can make sense of this user generated content and extract useful information

The content has the following characteristics:

- Diverse (i.e. multiple users, formats, extensions, domains, etc.)
- Chronological (i.e. diachronic analyses made possible as archived data is also digitalized, also data grows and evolves with societies everyday)

Case studies:

- Statistical analysis for evolution of language (meaning changes)
- Sentiment analysis for identifying changes in opinion

Language Evolution

Human language is subject to constant evolution driven by the need to reflect the ongoing changes in the world and to become a more efficient means of communication.

Areas of interest in language evolution include lexical change (focus here, see Jatowt and Duh [2014]), syntactic change, sound change, areal effects/borrowing, and mathematical models of evolution.

Diachronic linguistics (aka historical linguistics) investigates:

- The process of language development over time
- Why and how languages change and in which way these changes spread across spatio-temporal dimension

Word Meanings

Word meanings can be categorised as follows (Evert and Lenci [2009]):

- Meanings in the world: the meaning of *car* is the set of cars in this world (extension), or a function from possible words to the sets of cars in these worlds (e.g. formal semantics)
- Meanings in the head: the meaning of *car* is the concept CAR, as a mental representation of the category of cars (e.g. cognitive psychology)
- Meanings in the text: the meaning of *car* is an abstraction over the linguistic contexts in which the word *car* is used (e.g. distributional semantics)

The Distributional Hypothesis

Usage based perspective (Baroni and Boleda [2017], Evert and Lenci [2009]):

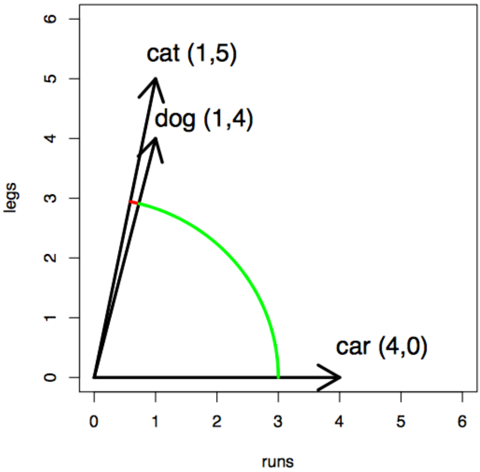
- The meaning of a word is the set of contexts in which it occurs in texts
- Important aspects of the meaning of a word are a function of (can be approximated by) the set of contexts in which it occurs in texts

"The meaning of a word is its use in the language" (Wittgenstein, 1953)
"You shall know a word by the company it keeps" (Firth, 1957)
"The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear." (Harris, 1954)

Distributional Semantic Models

- Computational models that build contextual semantic representations from corpus data
- Idea is to represent the meaning of words as vectors keeping track of the words' distributional history
- Semantic content for a word is represented by a vector, which is obtained through the statistical analysis of its linguistic contexts
  - count how many times each target word occurs in a certain context
  - build vectors out of (a function of) these context occurrence counts
- Similar words will have similar vectors
- Distributional similarity is interpreted as the semantic similarity

DSM Example



Visualisation of semantic space and semantic similarity as angle between vectors (Hamilton et al. [2016])

Parameters of DSMs (Evert and Lenci [2009])

When designing DSMs, it is essential to understand the effects of these parameters on the semantic properties of the models:

- Linguistic parameters
  - pre-processing and linguistic annotation - raw text, stemming, POS tagging and lemmatisation, (dependency) parsing, semantically relevant patterns
  - choice of context - document, sentence, window, dependency relations, etc.
- Mathematical parameters
  - context weighting - log-frequency, association scores, entropy, etc.
  - measuring distance - cosine similarity, Euclidean, Manhattan, etc
  - dimensionality reduction - feature selection, SVD projection (PCA), random indexing

Opinion Mining and Sentiment Analysis

Computational study of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, etc – expressed in text. See Liu [2011].

- Reviews, blogs, discussions, news, comments, feedback, etc

A popular research topic in NLP, text mining, and Web mining

Various applications include:

- Market intelligence for businesses
- Decision support for individuals to buy products/services
- Finding public opinions about issues, political candidates, etc.

Opinion Mining Problem Statement

Opinion mining is a very restricted NLP problem as the system only needs to understand (Cambria et al. [2013]):

- the positive or negative sentiments of each sentence
- the target entities or topics.

Two aspects of abstraction:

1. Structured definition of an opinion
   - document level, i.e., is this review + or -?
   - sentence level, i.e., is each sentence + or -?
   - entity and feature/aspect level
2. Opinion summarisation (opinions from many people)

Opinion Mining Problem Statement contd.

Opinion Definition

An opinion (or regular opinion) is simply a positive or negative sentiment, view, attitude, emotion, or appraisal about an entity or an aspect of the entity from an opinion holder.

Sentiment orientation of an opinion

Positive, negative, or neutral (no opinion). Also called opinion orientation, semantic orientation, sentiment polarity. Strength (e.g. very positive, $+0.8$, etc.) can matter in some cases.

Feature-based Sentiment Analysis Model

Model of an object

An object o is represented with a finite set of features, $F = f_1, f_2, ..., f_n$

Each feature $f_i \in F$ can be expressed with any one of a finite set of words or phrases $W_i = w_{i1}, w_{i2}, ..., w_{im}$, which are synonyms of the feature, or indicated by any one of a finite set of feature indicators $I_i = i_{i1}, i_{i2}, ..., i_{iq}$ of the feature.

Model of an opinionated document

A general opinionated document d contains opinions on a set of objects $o_1, o_2, ..., o_q$ from a set of opinion holders $h_1, h_2, ..., h_p$. The opinions on each object o_j are expressed on a subset F_j of features of o_j.

Feature-based Sentiment Analysis Model contd.

There are 5 essential components of a direct opinion, which can be defined as a quintuple

$$(o_j,\ f_{jk},\ oo_{ijkl},\ h_i,\ t_l)$$

- o_j is an object
- f_{jk} is a feature of the object o_j
- oo_{ijkl} is the opinion orientation and can be positive, negative, neutral, or more granular ratings.
- h_i is the opinion holder
- t_l is the time when the opinion is expressed

It is possible to add any number of other components to the tuple for more analysis (e.g., gender, age, web site, etc).

Sentiment Classification

In sentiment classification, opinion/sentiment words are more important, e.g., great, excellent, horrible, bad, worst, etc.

Document level [a]

[a] Perhaps the most widely studied problem where neutral is mostly ignored

Classify a whole opinion document (e.g., a review) based on the overall sentiment of the opinion holder.

- Classes: Positive, negative (possibly neutral)
- Assumption: The document is written by a single person and expresses opinion/sentiment on a single entity
- Goal: Discover $(\_, \_, oo, \_, \_)$, where o, f, h, and t are ignored
- Reviews usually satisfy the assumption. Almost all papers use reviews where *positive* is represented with 4 or 5 stars, and *negative* is represented with 1 or 2 stars.
- Two methods: Supervised and unsupervised

Supervised Document Level Classification

Existing supervised learning methods can be readily applied to sentiment classification, e.g., naïve Bayes, support vector machines (SVM), etc.

Key: feature engineering. Commonly used features include:

- Terms and their frequency: unigrams or n-grams and their frequency counts, word positions, TF-IDF[7]
- Part of speech tags: adjectives, adverbs
- Opinion words and phrases: words that are commonly used to express positive or negative sentiments. (e.g., beautiful, good, and amazing are positive, and bad, poor, and terrible are negative)
- Negation: negation words change the opinion orientation

Sentiment classification is highly sensitive to the domain from which the training data are extracted.

[7] Term Frequency and Inverse Document Frequency

Sentiment Classification contd.

Feature (aspect) level

Aspect level performs finer-grained analysis

- Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), directly looks at the opinion itself.
- Based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion).

Much of the research focuses on online reviews, which have usually known aspects. However, for blogs, news, discussions, etc., the problem is harder:

- Objects and aspects of the objects are unknown
- There may be many comparisons
- There may be irrelevant information

Sentiment Classification contd.

Feature (aspect) extraction

Finding aspects is a difficult task

- Explicit: Aspects explicitly mentioned as nouns or noun phrases in a sentence
- Implicit: Aspects not explicitly mentioned in a sentence but are implied, commonly using adjectives and adverbs. They require mapping either manually or computational methods (many papers on aspect extraction)

Once aspects are identified, find and group the synonyms into categories (e.g. power usage and battery life are the same) Methods include

- Lexical similarity based on WordNet
- Distributional information (surrounding words context)
- Syntactical constraints (sharing words, in the same sentence)

Sentiment Classification contd.

For each aspect, identify the sentiment

- Almost all approaches use opinion words and phrases, however context dependency is ignored (e.g. small and sucks are positive words for a vacuum cleaner)

Lexicon-based approaches need to deal with parsing of

- Simple, compound, comparative, conditional sentences, questions, etc.
- Negation (not), contrary (but), comparisons, etc.
- Context dependency

Sentiment Analysis over Time

Each word has certain sentiment value that can vary over time. In order to track sentiment changes in words, we first need a representation.

Word representation

Use of DSMs: each word is represented by its usage context in the n-gram datasets.

Semantic Change of Single Word

- construct vector representation of word context in each decade
- compare the context of the word in the last decade with the ones in the previous decades using (e.g.) cosine similarity calculation

Case Study: Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

Development of a methodology for quantifying semantic change by evaluating word embeddings against known historical changes Hamilton et al. [2016].

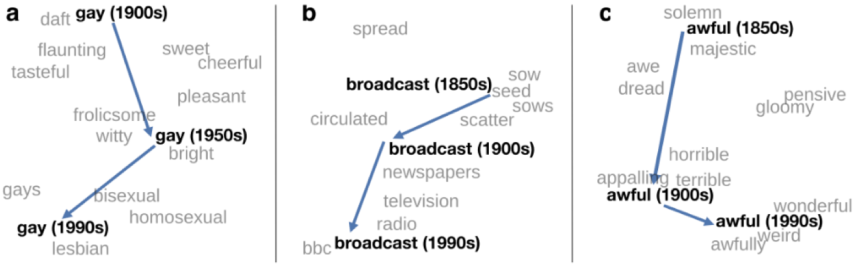


Figure 1: Visualization of semantic change in English Hamilton et al. [2016]

Case Study: Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change contd.

Data

- Large-scale cross-linguistic using 6 corpora spanning 200 years and 4 languages (English, German, French, and Chinese).

Findings

- Across languages, rates of semantic change obey a scaling relation in the form of power law.
- Frequent words change at slower rates while polysemous words change faster.

Conclusions

Ethical, legal and social implications (Schuller et al. [2016]):

- privacy and trust, traceability, explainability validity issues

Collected data is prone to bias

- Crowd-sourced annotation by large groups of individuals with often unknown reliability and high subjectivity

Next generation sentiment analysis and opinion mining systems need (Cambria et al. [2013]):

- broader and deeper common and commonsense knowledge bases
- more brain-inspired and psychologically motivated reasoning methods
- more efficient ways to bridge the gap between (unstructured) multi-modal information and (structured) machine-processable data

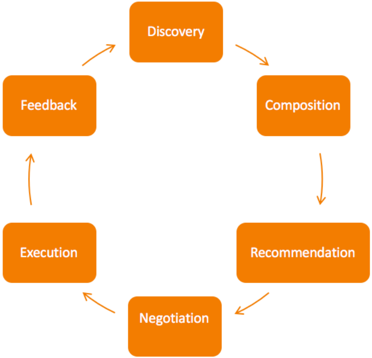
Part 5
From Behaviour to Agents

Agents vs. humans

- We saw how some ideas from ontologies can be applied to language
- Can we apply agent concepts to human behaviour in a similar way?
- Key questions:
  - How to structure real-world platforms applying these concepts?
  - Will human behaviour match the assumptions these methods make?
  - Can these techniques help us build more sustainable platforms?

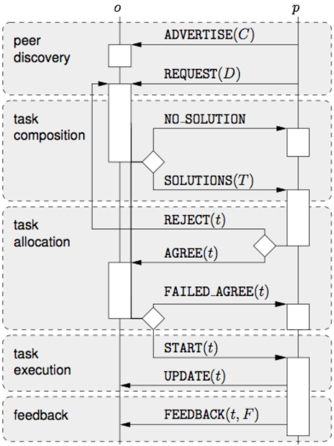
Social Orchestration

- Algorithms and architectures to enable organising human collaboration
- Emphasis on task composition, recommendation, and negotiation
- Challenges
  - How do we build models of human preferences?
  - How do we address diversity among them?

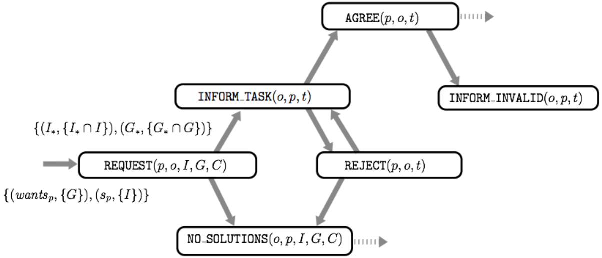


Interaction Models

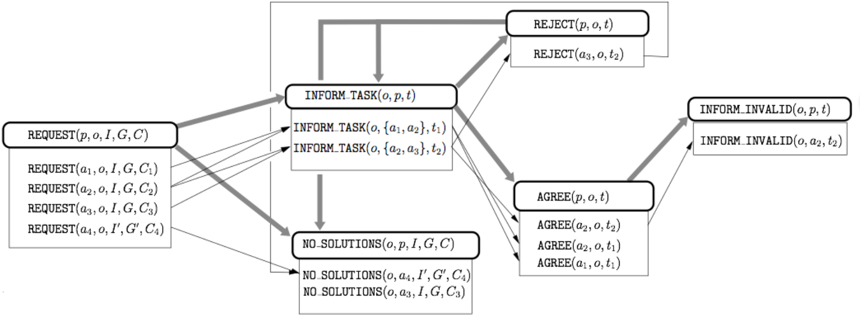
- First step: static social orchestration architecture
- Conceptualising interaction stages as a multiagent communication protocol
- Mapping this onto standard Web architectures (scalability, robustness, interoperablity)



Unfolding Agent Interactions



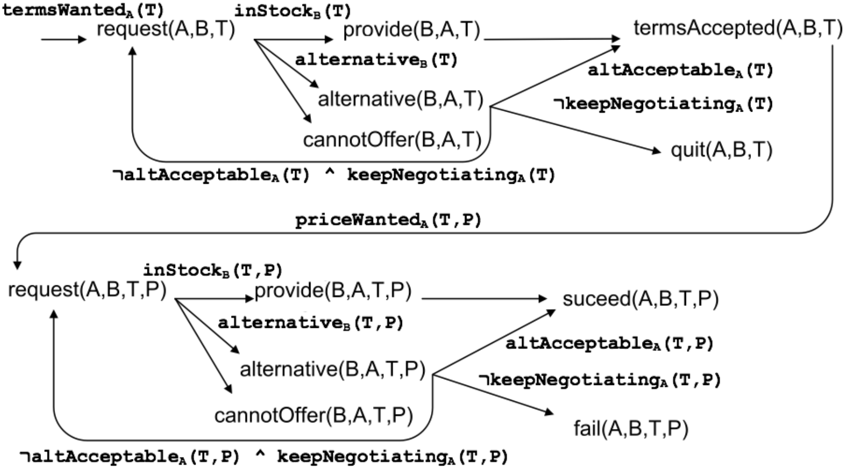
Unfolding Agent Interactions



Adaptive Social Orchestration

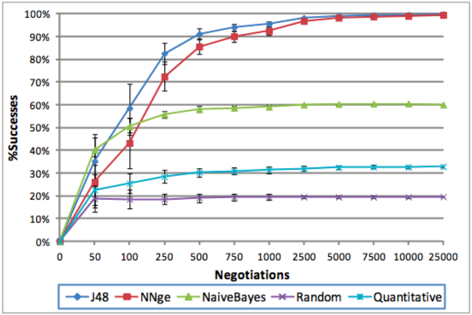
- For adaptive social orchestration we need to detect and exploit emergent patterns of behaviour
- Challenges
  - Large, fluctuating populations of agents
  - Unmanageable number of solutions
  - Individual and global outcomes correlated
- Three examples:
  - Mining agent conversations
  - Diversity-aware task recommendation
  - Coarse preference elicitation

Mining agent conversations



Qualitative mining of agent communication

```
persons = 2: F (158)
persons = 4: F (158)
persons = more
|   lug_boot = small
|   |   doors = 2: F (8)
|   |   doors = 3: F (7)
|   |   doors = 4: F (8)
|   |   doors = 5-more: T (105)
|   lug_boot = med
|   |   doors = 2: F (13)
|   |   doors = 3: F (8)
|   |   doors = 4: F (13)
|   |   doors = 5-more: T (120)
|   lug_boot = big: T (402)
```



Group task recommendation

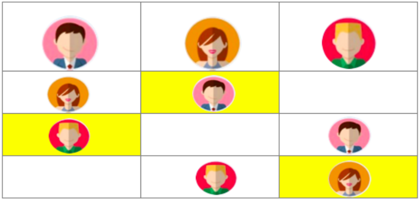
- Consider tasks involving group activity - want to compute optimal sets of solutions and influence user choice
  - E.g. sharing economy, smart cities, autonomous vehicles
- Shared resources and skills, varying individual user needs and preferences, global platform objectives
- Existing commercial apps offer ad hoc solutions
  - require lots of data/human input, no transfer/reuse

Group task recommendation

- Diverse set of (incompatible?) preferences
- Global utility depends on social welfare and task completion

$$U_s = \sum_{i \in I} u_i + \sum_{i \in I} \sum_{j \in J} x_{i,j}$$

- Hard to bound losses when users make individual choices
- Traditional mechanism design not applicable, we cannot enumerate solutions



Optimisation procedure

MIP* $\rightarrow V^*$	**Objective**	$max_{a \in A} U_s(a)$
	Constraints	Hard feasibility constraints
MIP$^{\text{first}}$ $\rightarrow a \in R$	**Objective**	$min_{a \in A} \sum_{i \in I} \sum_{i' \in I \mid i' > i} \lvert u_i(a) - u_{i'}(a) \rvert$
	Constraints	MIP* $U_s(a) \cdot h \geq V^*$
MIP$^{\text{others}}$ $\rightarrow a' \in R$	**Objective**	$min_{a' \in A} \sum_{i \in I} \lvert u_i(a) - u_i(a') \rvert$
	Constraints	MIP$^{\text{first}}$ $a' \notin R$

Taxation model

MIP*

$\rightarrow V^*$

Sponsored Solution

MIP$^{\text{first}}$

$\rightarrow a \in R$

MIP$^{\text{others}}$

$\rightarrow a' \in R$

Objective

$$min \sum_{i \in I} |u_i(a) - u_i(a') + \tau_i(a')| + M\Big(\sum_{i \in I} (u_i(a) + \epsilon - u_i(a) + \tau_i(a'))\Big)$$

Constraints MIP$^{\text{first}}$ $a' \notin R$

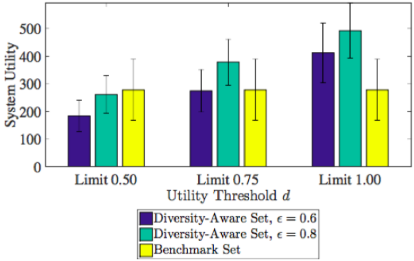
Noiseless and Constant Noise Models

$$u_i(a) + \epsilon \geq u_i(a') - \tau_i(a')$$

Logit Model (also goes into objective function)

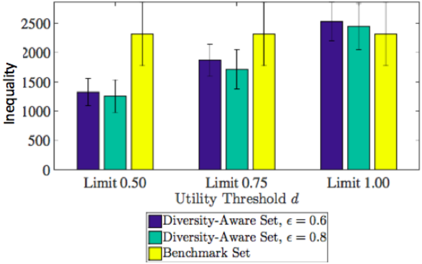
$$\frac{u_i(a)}{\Big(\sum_{a'' \in R} (u_i(a'') - \tau_i(a'')) + u_i(a') - \tau_i(a')\Big)} \geq \psi$$

Results



System Utility
0
100
200
300
400
500
Limit 0.50
Limit 0.75
Limit 1.00
Utility Threshold d
Diversity-Aware Set, $\epsilon = 0.6$
Diversity-Aware Set, $\epsilon = 0.8$
Benchmark Set

(a)

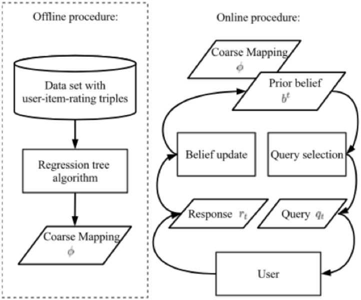


Inequality
0
500
1000
1500
2000
2500
Limit 0.50
Limit 0.75
Limit 1.00
Utility Threshold d
Diversity-Aware Set, $\epsilon = 0.6$
Diversity-Aware Set, $\epsilon = 0.8$
Benchmark Set

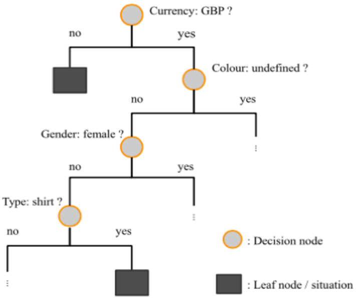
(b)

Coarse preference elicitation

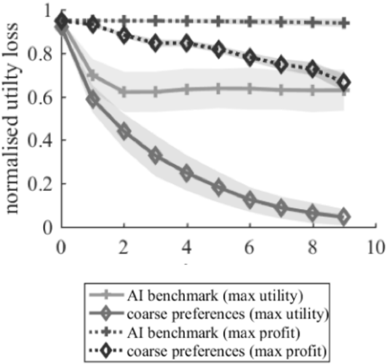
- Assume we have categories of items that are sufficient to express users' preferences
- Want to maximise value of item presented to user (given valuation function, e.g. myopic)
- Normal Bayesian update of user's utility function becomes simpler
- But we also learn more accurate models faster



Results



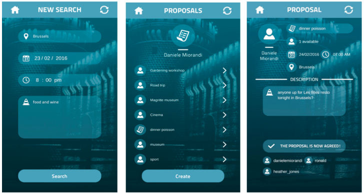
Currency: GBP ?
no
yes
Colour: undefined ?
no
yes
Gender: female ?
no
yes
Type: shirt ?
no
yes
: Decision node
: Leaf node / situation



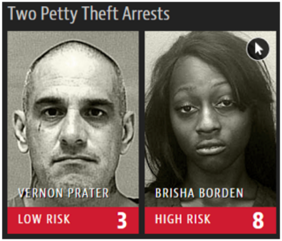
normalised utilty loss
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
AI benchmark (max utility)
coarse preferences (max utility)
AI benchmark (max profit)
coarse preferences (max profit)

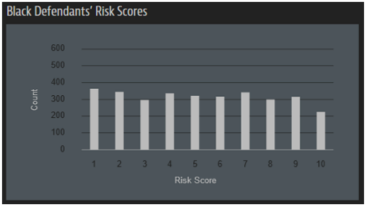
First steps toward a grand vision

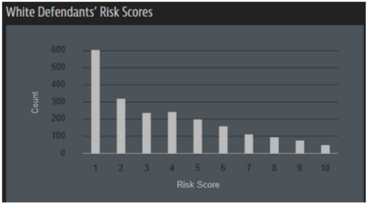
- AskAndShare: An app for doing things together without knowing the what/who/how
- Combines human computation with machine-driven recommendation
- Enabling general collective problem solving in human-AI systems



Ethics of data-driven systems



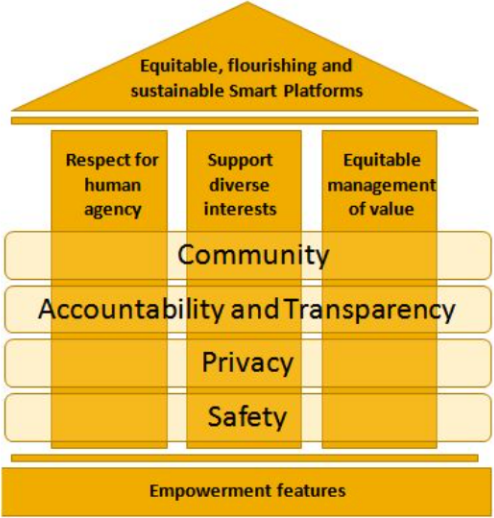




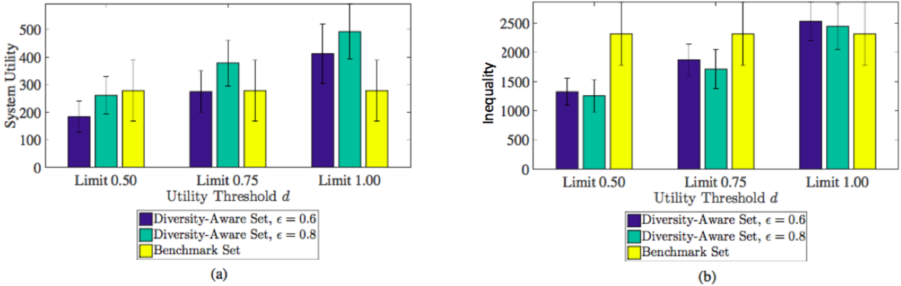
Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

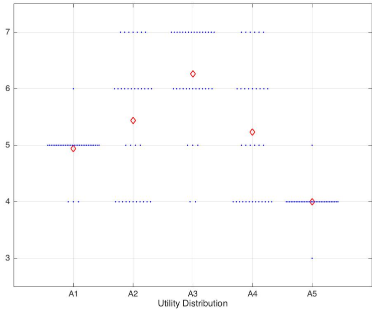
Social Charter for Smart Platforms

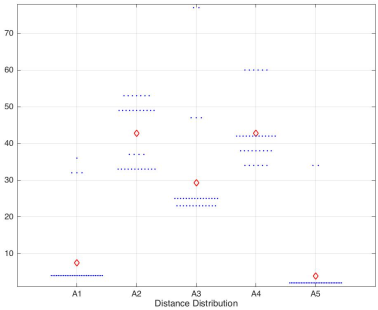


Results


System Utility
500
400
300
200
100
0
Limit 0.50
Limit 0.75
Limit 1.00
Utility Threshold d
Diversity-Aware Set, $\epsilon = 0.6$
Diversity-Aware Set, $\epsilon = 0.8$
Benchmark Set
(a)
Inequality
2500
2000
1500
1000
500
0
Limit 0.50
Limit 0.75
Limit 1.00
Utility Threshold d
Diversity-Aware Set, $\epsilon = 0.6$
Diversity-Aware Set, $\epsilon = 0.8$
Benchmark Set
(b)

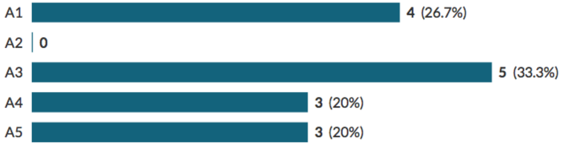
What is fair?


7
6
5
4
3
A1
A2
A3
A4
A5
Utility Distribution

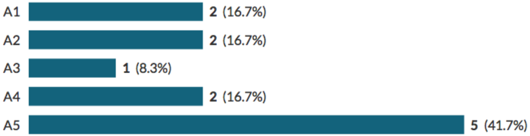

70
60
50
40
30
20
10
A1
A2
A3
A4
A5
Distance Distribution

What is fair?

Most preferred algorithm:



Least preferred algorithm:

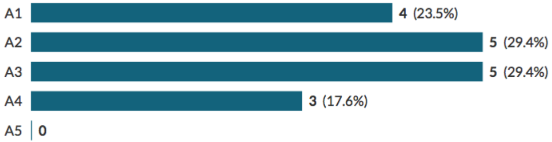


What is fair?

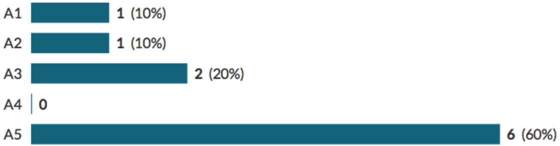
- A1 minimises differences among individuals while guaranteeing at least 70% of maximum possible social welfare
- A2 maximises the minimum individual utility while guaranteeing at least 70% of the maximum possible social welfare
- A3 maximises total social welfare
- A4 maximises the minimum individual utility
- A5 minimises total difference among individual utilities

What is fair?

Most preferred algorithm (after seeing description):



Least preferred algorithm (after seeing description):



Ethics – challenges

- Behaviour of algorithms depends on input data supplied
- Mathematical details of algorithms hard to grasp
- Multiple metrics relevant, some of which are incompatible
- Combining metrics creates representation and complexity problems
- Multiple stakeholders with different priorities regarding fairness
- Collective behaviour raises issues of shared responsibility

Ethics – challenges

- Cognitive difficulty of understanding provenance of results
- Algorithm details are commercially sensitive business assets
- Platforms do not necessarily have clear boundaries
- Legal requirements hard to align/enforce internationally
- Individual sectors dominated by oligopolists/monopolists
- Governance frameworks and regulatory institutions missing

Summary

New forms of **collective people-machine intelligence** are emerging on the Web. Complex dynamics of adaptation govern **evolution of interaction and meaning** in these systems. **Agent-based and semantic techniques** help understand and improve such platforms. Combining these methods with **data-driven technologies** is necessary to capture operation *in situ*. **Socio-technical nature** of systems brings challenges of scale and responsibility.

We are still far from having a science of these systems, only baby steps so far. But if we get this right, such systems can help solve the big challenges facing humanity.

Thank you for your attention!

References I

Marco Baroni and Gemma Boleda. Distributional semantics, 2017. URL https://www.cs.utexas.edu/~mooney/cs388/slides/dist-sem-intro-NLP-class-UT.pdf.

Lorenz Bühmann, Daniel Fleischhacker, Jens Lehmann, Andre Melo, and Johanna Völker. Inductive lexical learning of class expressions. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 42–53. Springer, 2014.

Erik Cambria, Björn Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2):15–21, 2013.

Philipp Cimiano and Johanna Völker. text2onto. In *International Conference on Application of Natural Language to Information Systems*, pages 227–238. Springer, 2005.

Philipp Cimiano, Christina Unger, and John McCrae. Ontology-based interpretation of natural language. *Synthesis Lectures on Human Language Technologies*, 7(2):1–178, 2014.

James R Curran, Stephen Clark, and Johan Bos. Linguistically motivated large-scale nlp with c&c and boxer. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 33–36. Association for Computational Linguistics, 2007.

S Evert and A Lenci. Foundations of distributional semantic models. ESSLLI 2009, 2009.

Ethan Fast and Eric Horvitz. Long-term trends in the public perception of artificial intelligence. *arXiv preprint arXiv:1609.04904*, 2016.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*, 2016.

Amelia Heathman. This ai found trends hidden in british history for more than 150 years, 2017. URL http://www.wired.co.uk/article/british-history-big-data-ai.

Adam Jatowt and Kevin Duh. A framework for analyzing semantic change of words across time. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 229–238. IEEE Press, 2014.

Hans Kamp and Uwe Reyle. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*, volume 42. Springer Science & Business Media, 2013.

References II

Bing Liu. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition*, pages 627–666. Chapman and Hall/CRC, 2010.

Bing Liu. Sentiment analysis and opinion mining, 2011. URL `https://www.cs.uic.edu/~liub/FBS/Sentiment-Analysis-tutorial-AAAI-2011.pdf`.

Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.

John P McCrae, Christian Chiarcos, Francis Bond, Philipp Cimiano, Thierry Declerck, Gerard de Melo, Jorge Gracia, Sebastian Hellmann, Bettina Klimek, Steven Moran, et al. The open linguistics working group: Developing the linguistic linked open data cloud. In *Proc. 10th Language Resources and Evaluation Conference (LREC 2016), Portoroz, Slovenia*, 2016.

Alina Petrova, Yue Ma, George Tsatsaronis, Maria Kissa, Felix Distel, Franz Baader, and Michael Schroeder. Formalizing biomedical concepts from textual definitions. *Journal of biomedical semantics*, 6(1):1, 2015.

Valentina Presutti, Giorgia Lodi, Andrea Nuzzolese, Aldo Gangemi, Silvio Peroni, and Luigi Asprino. The role of ontology design patterns in linked data projects. In *Conceptual Modeling: 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings 35*, pages 113–121. Springer, 2016.

Björn Schuller, Jean-Gabriel Ganascia, and Laurence Devillers. Multimodal sentiment analysis in the wild: Ethical considerations on data collection, annotation, and exploitation. In *Actes du Workshop on Ethics In Corpus Collection, Annotation & Application (ETHI-CA2), LREC, Portoroz, Slovénie*, 2016.

Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

Johanna Völker, Daniel Fleischhacker, and Heiner Stuckenschmidt. Automatic acquisition of class disjointness. *Web Semantics: Science, Services and Agents on the World Wide Web*, 35:124–139, 2015.