A Game-Theoretic Approach to Word Sense Disambiguation

Slides

Slide 1: Paper title, authors, our names, etc

Slide 2: Paper intro - Abstract
- **Classic game theory VS Evolutionary game theory**
- **Static** strategies (**CGT**) VS **Dynamic** strategies (**EGT**)
- **Evolutionary game theory** (**EGT**) is the application of game theory to evolving populations in biology. It defines a framework of contests, strategies, and analytics into which Darwinian competition can be modelled.
- Natural Selection
- payoffs -> reproductive success

Slide 3 : Related work - Problems

Slide 4 : Solution - Proposed system
- It assumes that there is a **population** of individuals, represented by all the **senses** of the words to be disambiguated, and that there is a **selection process**, which selects the **best candidates** in the population. The selection process is defined as a **sense similarity function**, which **gives a higher score to candidates with specific features**, increasing their fitness to the detriment of the other population members. This process is repeated until the fitness level of the population regularizes and at the end the candidates with higher fitness are selected as solutions of the problem.
- our approach is defined in terms of **evolutionary game theory**.

Slide ? : Strengths / Contribution
-

Slides 5++ : Data modelling & Math
(p.43)
- Brief info on dataset
- Περιγραφή του βήματος - μαθηματική ανάλυση (σημειογραφία / τύποι / πίνακες / τεχνική εφαρμογή / γραφήματα-περιγραφή / φυσική σημασία)
- Παράλληλη αναφορά στο παράδειγμα με το river bank (φλεξ με prisoner's dilemma - game theory έννοιες)

Slide 6: Step 2
- Compute from I the word similarity matrix W in which are stored the pairwise similarities among each word with the others and represents the players' interactions
- graph construction (5.1.1)
- co-occurrence graph of example (Figure 4a)
    - graph: geometry of the data
    - nodes: words

- weighted edges: m-dice similarity(association measure) between words i,j

Slide 7: Step 3
- increase the weights between two words that share a proximity relation
- n-gram, combination of similarity & proximity (n-nearest left & right neighbours)

Slide 8: Step 4
- extract from **I** the list C of all the possible senses that represent the strategy space of the system
-

Slide 9: Step 5
- assign for each word in I a probability distribution over the senses in C creating for each player a probability distribution over the possible strategies

Slide 10: Step 6
- compute the sense similarity matrix Z among each pair of senses in C, which is then used to compute the partial payoff matrices of each game

Slide 11: Step 7
- Apply the replicator dynamics equation in order to compute the Nash equilibria of the games
- If each player has chosen [a strategy]—an action plan choosing his own actions based on what it has seen happen so far in the game—and no player can increase his own expected payoff by changing his strategy while the other players keep theirs unchanged, then the current set of strategy choices constitutes a **Nash equilibrium**.
- **Nash equilibria** correspond to **consistent word-sense assignments**!
- There are several reasons for the **prominent role of replicator dynamics:** Firstly,(1) is relatively simple and mathematically well understood (Hofbauer and Sigmund2003). It is equivalent to the famous Lotka–Volterra model in population ecology(Hofbauer1981). Additionally, there are beautiful connections between replicator dynamics and the concepts of classical game theory (Fudenberg and Tirole1991;We i bu l l1995). For example, **if strategy i is dominated**, meaning that **there is another strategy which is always better**, then **replicator dynamics will lead to the extinction of i** (through computations - **equation 6**)

Slide 12: Step 8
- assign to each word $i \in I$ a strategy $s \in C$

Slides 13-14: Parameter Tuning & Experiments
- (WordNet+BabelNet + algorithm comparison + graphs)
- Τελικές παράμετροι - με βάση προηγούμενα πειράματα κλπ κλπ

        - mdice (association measure - weighted edges of graph W - word similarity)
        - tf-idf (payoffs)
        - n=5 (n-gram)
        - p=0.4 (semi-supervised learning)
    - F-score
        - **precision:** Precision represents the proportion of items – in this case, entities – that the system returns which are accurately correct. It rewards careful selection, and punishes over-zealous systems that return too many results: to achieve high preci-sion, one should discard anything that might not be correct.False positives – spurious entities – decrease precision.
        - **recall:** Recall indicates how much of all items that should have been found, were found. This metric rewards comprehen-siveness: to achieve high recall, it is better to include en-tities that one is uncertain about. False negatives – missed entities – lead to low recall. It balances out precision.
        - **? αντιστοιχία με το παράδειγμα ?**

Slide 15: Conclusion & Strengths

Slide 16: Smart closing

- n-gram / co-occurrence graph με αμφίσημες λέξεις / βάρη, κλπ

Notes

- Φυσική σημασία **equation 6** : η πιθανότητα (t+1) να επιλεγεί η έννοια h από τη λέξη i εξαρτάται από την πιθανότητα (t) και το άθροισμα των προτιμήσεων των υπόλοιπων λέξεων ως προς αυτή την έννοια (utility - payoff) διαιρούμενη από το μέσο όρο (average) των προτιμήσεων όλων των λέξεων για όλες τις έννοιες.
- **Relative fitness:** https://en.wikipedia.org/wiki/Fitness_(biology): Relative fitnesses only indicate the change in prevalence of different genotypes relative to each other, and so only their values relative to each other are important; relative fitnesses can be any nonnegative number, including 0.
- The **expected utility hypothesis** is a popular concept in economics, game theory and decision theory that serves as a reference guide for judging decisions involving uncertainty.[1] The theory recommends which option a rational individual should choose in a complex situation, based on his tolerance for risk and personal preferences. The expected utility of an agent's risky decision is the mathematical expectation of his utility from different outcomes given their probabilities
-

Definitions

- **Altruism**: When one organism reduces its own fitness to benefit the fitness of another organism.
- **Evolutionary stable strategy**: A behavioral strategy (phenotype) if adopted by all individuals in a population that cannot be replaced or invaded by a different strategy through natural selection.
- **Game theory**: A mathematical approach to understanding the outcomes of interactions between two or more individuals when benefits and costs of the interactions depend on the strategies of each individual.
- **Inclusive fitness**: The fitness of a gene as measured by the fitness of the individual possessing the gene and the fitness of the individual's relatives bearing the same gene, identical by descent.
- **Mutualism**: A relationship between two individuals from different species that benefits each individual involved in the interaction.
- **Phenotype**: The physical, physiological, behavioral and other traits expressed by an individual.

Cheat Sheet

1. The WSD problem can be formulated in game-theoretic terms modeling:

•the players of the games as the words to be disambiguated.
•the strategies of the games as the senses of each word.
•the payoff matrices of each game as a sense similarity function.
•the interactions among the players as a weighted graph.
Nash equilibria correspond to consistent word-sense assignments!
•Word-level similarities: proportional to strength of co-occurrence between words
Sense-level similarities: computed using WordNet / BabelNet ontologies

Game theory offers a principled and viable solution to context-aware pattern recognition problems, based on the idea of dynamical competition among hypotheses driven by payoff functions.
Distinguishing features:
•No restriction imposed on similarity/payoff function (unlike, e.g., spectral methods)
•Shifts the emphasis from *optima* of objective functions to *equilibria* of dynamical systems.
On-going work:
•Learning payoff functions from data (Pelillo and Refice, 1994)
•Combining Hume-Nash machines with deep neural networks
•Applying them to computer vision problems such as scene parsing, object recognition, video analysis

---