A Game-Theoretic Approach to Word Sense Disambiguation

Slides

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

Slide 2: Introduction

1. **a new model for word sense disambiguation formulated in terms of evolutionary game theory**
2. **Word Sense Disambiguation (WSD)**: the task of identifying the intended meaning of a word based on the context in which it appears
3. **Example ->** There is a financial institution near the river bank

- **a new model for word sense disambiguation formulated in terms of evolutionary game theory**
- Word Sense Disambiguation (WSD) is the task of identifying the intended meaning of a word based on the context in which it appears (eg. There is a financial institution near the river bank)

Spoken text:
Οι συγγραφείς προσεγγίζουν το WSD μέσω του game theory.

Slide 2b: Game Theory

1. A mathematical approach to understand the outcomes of interactions between two or more individuals when benefits and costs of the interactions depend on the strategies of each individual.(+ definitions)
   a. **players** I (= {1, .., n})
   b. **pure strategies** for each player Si (= {s1, ..., sm})
   c. **mixed strategies x :** A mixed strategy set can be defined as a vector x=(x1,...,xm), where m is the number of pure strategies and each component xh denotes the probability that player i chooses its hth pure strategy
   d. **utility** function ui (: S1 × ... × Sn → R) : associates strategies to payoffs

2. **Classical game theory VS Evolutionary game theory**
   a. **Static** strategies (**CGT**) VS **Dynamic** strategies (**EGT**)
   b. **Evolutionary game theory** (EGT)**:**
      i. overcoming some limitations of traditional game theory

        ii. Evolutionary game theory differs from classical game theory in focusing more on the dynamics of strategy change (επανάληψη).

    c. **Nash equilibria**: strategy profiles in which each strategy is a best response to the strategy of the co-player and no player has the incentive to deviate from their decision (no way to do better)
    d. **Payoff** matrix:

3. WSD problem modelling in game theoretical terms
    a. players -> words
    b. strategies -> senses (evolving population)
    c. payoff matrices -> sense similarity
    d. interactions -> weighted graph
    e. 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.

- **Game theory** provides predictive power in interactive decision situations. It was intro- duced by Von Neumann and Morgenstern (1944) in order to develop a mathematical framework able to model the essentials of decision-making in interactive situations. In its normal form representation (which is the one we use in this article) it consists of a finite set of players **I** = {1, .., n}, a set of pure strategies for each player Si = {s1, ..., sm}, and a utility function ui : S1 × ... × Sn → R, which associates strategies to payoffs. Each player can adopt a strategy in order to play a game; and the utility function depends on the combination of strategies played at the same time by the players involved in the game, not just on the strategy chosen by a single player.
- **Classical game theory VS Evolutionary game theory**
- A game theoretic framework can be considered as a solid tool in decision-making situations because a fundamental theorem by Nash (1951) states that any normal-form game has at least one mixed Nash equilibrium, which can be used as the solution of the decision problem.
- **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
- supervised and knowledge-based. Supervised algorithms learn, from sense-labeled corpora, a computational model of the words of interest. Then, the obtained model is used to classify new instances of the same words.

Knowledge-based algorithms perform the disambiguation task by using an existing lexical knowledge base, which usually is structured as a semantic network. Then, these approaches use graph algorithms to disambiguate the words of interest, based on the relations that these words' senses have in the network (Pilehvar and Navigli 2014).

- 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
-

Slide 3 : Background

- Multiple NLP applications can **benefit** from the disambiguation of ambiguous words, as a preliminary process; otherwise they remain on the surface of the word, compromising the coherence of the data to be analyzed
- **supervised :** model learns from sense-labeled corpora and then is obtained to classify new instances
- **knowledge-based :** using an existing lexical knowledge base, use graph algorithms to disambiguate the words based on the relations that these words' senses have in the network
- **unsupervised :** learns patterns from untagged data
- **semi-supervised :** small amount of labelled & large amount of unlabelled data
- **heuristics :** practical (not necessarily optimal) solution to a problem
- **graph-based :** model the relations among words and senses in a text with graphs, representing words and senses as nodes and the relations among them as edges

Slide 4 : Proposed system

1. Our approach is defined in terms of **evolutionary game theory**.
2. WSD task
    a. **sense-labeling task** (sense assignment to word)
    b. **constraint satisfaction problem**
3. WSD problem modelling in game theoretical terms
    - **players** -> words
    - **strategies** -> senses (evolving population)

    - **payoff** matrices -> sense similarity
    - **interactions** -> 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
    - 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.
4. this approach ensures that the final labeling of the data is consistent and that the solution of the problem is always found. In fact, our system **always converges to the nearest Nash equilibrium** from which the dynamics have been started . **TODO - always converges**
5. This approach gives us the possibility not only to exploit the **contextual** information of a word but also to find **the most appropriate sense association for the target word and the words in its context.**
6. **versatile approach:** method is adaptive to different scenarios and to different tasks, and it is possible to use it as **unsupervised** or **semi-supervised** (new semi-supervised version of the approach, which can exploit the evidence from sense tagged corpora or the most frequent sense heuristic and does not require labeled nodes to propagate the labeling information.).

Slides 5++ : Data modelling & Math
(p.43)
- Brief info on dataset (WordNet, corpus used, etc)
- 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

1. 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
2. **similarity matrix W**

- 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
- WordNet

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
- **Sense-level similarities**: computed using WordNet / BabelNet ontologies

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

- We **tested** our approach on different data sets from WSD and **entity-linking tasks** in order to find the **similarity measures that perform better**, and **evaluated** our **approach** against **unsupervised, semi-supervised, and supervised state-of-the-art systems**.
- (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 (F1) - μαθηματικός τυπος (αρμονικός μεσος)
    - **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 precision, 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 entities that one is uncertain about. False negatives – missed entities – lead to low recall. It balances out precision.
    - **Precision:** το ποσοστό των οντοτήτων που κατηγοριοποιήθηκαν (αποσαφηνίστηκαν) σωστά προς το γενικό σύνολο οντοτήτων που κατηγοριοποιήθηκε
    - **Recall:** το ποσοστό των οντοτήτων που κατηγοριοποιήθηκαν σωστά προς το σύνολο των οντ
    - In a classification task, the precision for a class is the number of true positives (i.e. the number of items correctly labelled as belonging to the positive class) divided by the total number of elements labelled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labelled as belonging to the class). Recall in this context is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labelled as belonging to the positive class but should have been).
    - Precision is the estimated probability that a document randomly selected from the pool of retrieved documents is relevant.
    - Recall is the estimated probability that a document randomly selected from the pool of relevant documents is retrieved.
    -
    -
- The results of this evaluation show that **our method performs well and can be considered as a valid alternative to current models**.
- **WordNet** as knowledge-base
- **BabelNet** as knowledge-base

Slide 15: Key Findings
(all evaluation results, bullet points)

Slide 16: Conclusion & Strengths

1. **The** exploitation of the **contextual** information of a word along with the fact that we also find **the most appropriate sense association for the target word and the words in its context, is the most important contribution of our work,** which distinguishes it from existing WSD algorithms. In fact, in some cases using only contextual information without the imposition of constraints can lead to inconsistencies in the assignment of senses to related words.
2.

Slide 17: 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.
- A **fitness function** is a particular type of objective function that is used to summarise, as a single figure of merit, how close a given design solution is to achieving the set aims. Fitness functions are used in genetic programming and genetic algorithms to guide simulations towards optimal design solutions.
- 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