Things that keep me awake at nights (well, not exactly but anyway...)

Stergios Chatzikyriakidis

February 7th, 2017

centre for
linguistic theory
and studies in probability

First shot: In defence of lost causes? Logic in Computational Linguistics

- The Machine Learning/Deep Learning revolution (well you are free not to accept this term) in CL
- Limitations of symbolic approaches (costly, they break easily)
  - What is the role to be played by symbolic approaches in practical NLP applications in the future?

CLASP centre for linguistic theory and studies in probability

Logic and vector space semantics, logic and probability, logic and ML/DL

- Attempts to combine vector space semantics with logic exist (e.g. work by Baroni et al., Sadrazdeh et al. etc)
  - Some things seem to be incompatible, e.g. negation seems rather difficult
- Same with probabilistic logics (Erk, Cooper, Lappin et al.)
- Not clear how logic and DL can fit together is some meaningful way
  - Learning logical systems? (work by Socher et al. on using NNs to capture logical inference, set inclusion relations etc.)
- First conference on logic and machine learning at university of Gothenburg
  - Consider submitting! [Link]

CLASP centre for linguistic theory and studies in probability

Natural Language Inference

- What is Natural Language Inference
  - Basically something like this: how much of human reasoning should a CL system be able to capture
  - FraCaS, PASCAL textual entailment and recently SNLI: all seem to concentrate on aspects of what NLI is
    - Also, all of three seem to be designed, having (implicitly possibly) in mind the standard line of research at that given period
  - Gradience and inference

CLASP centre for linguistic theory and studies in probability

Natural Language Inference

- What counts as an NLI?
  - Logical Entailment
    - ⋆ The FraCaS test (Cooper et al. 1996) suite provides a collection of mostly logical entailments. Categorization is done according to semantic category
    - ⋆ Three way classification of 346 inference problems: YES (the conclusion follows), NO (the negation of the conclusion follows) and UNK (none of the two follow)

CLASP centre for linguistic theory and studies in probability

Natural Language Inference

- What counts as an NLI?
  - Logical Entailment
    - ★ The FraCaS test (Cooper et al. 1996) suite provides a collection of mostly logical entailments. Categorization is done according to semantic category
    - ★ Three way classification of 346 inference problems: YES (the conclusion follows), NO (the negation of the conclusion follows) and UNK (none of the two follow)

Natural Language Inference

(1) A Swede won the Nobel Prize.
Every Swede is Scandinavian.
Did a Scandinavian win the Nobel prize? [Yes, FraCas 049]

Natural Language Inference

(3) A Swede won the Nobel Prize.
Every Swede is Scandinavian.
Did a Scandinavian win the Nobel prize? [Yes, FraCas 049]

(4) No delegate finished the report on time..
Did any Scandinavian delegate finish the report on time? [No, FraCaS 070]

Natural Language Inference

(5) A Scandinavian won the Nobel Prize.
Every Swede is Scandinavian.
Did a Swede win the Nobel prize? [UNK, FraCaS 065]

CLASP centre for linguistic theory and studies in probability

Natural Language Inference

(9) A Scandinavian won the Nobel Prize.
Every Swede is Scandinavian.
Did a Swede win the Nobel prize? [UNK, FraCaS 065]

- Other typical examples

(10) Either Smith, Jones or Anderson signed the contract. Did John sign the contract? [UNK] (plurals, FraCaS 083)

CLASP centre for linguistic theory and studies in probability

Natural Language Inference

(13) A Scandinavian won the Nobel Prize.
Every Swede is Scandinavian.
Did a Swede win the Nobel prize? [UNK, FraCaS 065]

- Other typical examples

  (14) Either Smith, Jones or Anderson signed the contract. Did John sign the contract? [UNK] (plurals, FraCaS 083)

  (15) Dumbo is a large animal. Is Dumbo a small animal? [NO] (adjectives, FraCaS 205)

Natural Language Inference

(17) A Scandinavian won the Nobel Prize.
Every Swede is Scandinavian.
Did a Swede win the Nobel prize? [UNK, FraCaS 065]

- Other typical examples

(18) Either Smith, Jones or Anderson signed the contract. Did John sign the contract? [UNK] (plurals, FraCaS 083)

(19) Dumbo is a large animal. Is Dumbo a small animal? [NO] (adjectives, FraCaS 205)

(20) Smith believed that ITEL had won the contract in 1992. Did ITEL win the contract in 1992? [UNK] (Attitudes, FraCaS 334)

CLASP centre for linguistic theory and studies in probability

Natural Language Inference

(21) A Scandinavian won the Nobel Prize.
Every Swede is Scandinavian.
Did a Swede win the Nobel prize? [UNK, FraCaS 065]

- Other typical examples

  (22) Either Smith, Jones or Anderson signed the contract. Did John sign the contract? [UNK] (plurals, FraCaS 083)

  (23) Dumbo is a large animal. Is Dumbo a small animal? [NO] (adjectives, FraCaS 205)

  (24) Smith believed that ITEL had won the contract in 1992. Did ITEL win the contract in 1992? [UNK] (Attitudes, FraCaS 334)

CLASP centre for linguistic theory and studies in probability

The FraCaS: Brief evaluation

- Unnatural data
  - Highly skilled formal semanticists provided types of examples that an NLI system should be able to capture

CLASP centre for linguistic theory and studies in probability

The FraCaS: Brief evaluation

- Unnatural data
  - Highly skilled formal semanticists provided types of examples that an NLI system should be able to capture
  - Small dataset for today's world
- However the FraCaS is a useful resource since it contains targeted examples according to the linguistic phenomena involved

CLASP centre for linguistic theory and studies in probability

The FraCaS: Brief evaluation

- Unnatural data
  - Highly skilled formal semanticists provided types of examples that an NLI system should be able to capture
  - Small dataset for today's world
- However the FraCaS is a useful resource since it contains targeted examples according to the linguistic phenomena involved
  - Involves cases of very fine-grained inference that the newer platforms do not have (e.g. reasoning with elliptical fragments, aspectual inference etc.)

CLASP centre for linguistic theory and studies in probability

The RTE platforms

- Platform appeared in 2005 (Dagan et al. 2006)

CLASP centre for linguistic theory and studies in probability

The RTE platforms

- Platform appeared in 2005 (Dagan et al. 2006)
  - New version every year until 2011 (RTE7)

CLASP centre for linguistic theory and studies in probability

The RTE platforms

- Platform appeared in 2005 (Dagan et al. 2006)
  - New version every year until 2011 (RTE7)
  - RTE concentrates on real rather than constructed data

The RTE platforms

- Platform appeared in 2005 (Dagan et al. 2006)
  - New version every year until 2011 (RTE7)
  - RTE concentrates on real rather than constructed data
  - A natural piece of text is taken as a premise and then a hypothesis is constructed out of it

CLASP centre for linguistic theory and studies in probability

The RTE platforms

- Platform appeared in 2005 (Dagan et al. 2006)
  - New version every year until 2011 (RTE7)
  - RTE concentrates on real rather than constructed data
  - A natural piece of text is taken as a premise and then a hypothesis is constructed out of it
    - ⋆ The first RTEs had two way entailment (entailment, non-entailment), three way classification was added in the later suites

CLASP centre for linguistic theory and studies in probability

The RTE platforms

- Platform appeared in 2005 (Dagan et al. 2006)
  - New version every year until 2011 (RTE7)
  - RTE concentrates on real rather than constructed data
  - A natural piece of text is taken as a premise and then a hypothesis is constructed out of it
    - The first RTEs had two way entailment (entailment, non-entailment), three way classification was added in the later suites

CLASP centre for linguistic theory and studies in probability

Some RTE examples

P Budapest again became the focus of national political drama in the late 1980s, when Hungary led the reform movement in eastern Europe that broke the communist monopoly on political power and ushered in the possibility of multiparty politics.

H In the late 1980s Budapest became the center of the reform movement. [follows, RTE702]

CLASP centre for linguistic theory and studies in probability

Some RTE examples

P Budapest again became the focus of national political drama in the late 1980s, when Hungary led the reform movement in eastern Europe that broke the communist monopoly on political power and ushered in the possibility of multiparty politics.

H In the late 1980s Budapest became the center of the reform movement. [follows, RTE702]

CLASP centre for linguistic theory and studies in probability

Some RTE examples

P Budapest again became the focus of national political drama in the late 1980s, when Hungary led the reform movement in eastern Europe that broke the communist monopoly on political power and ushered in the possibility of multiparty politics.

H In the late 1980s Budapest became the center of the reform movement. [follows, RTE702]

P Like the United States, U.N. officials are also dismayed that Aristide killed a conference called by Prime Minister Robert Malval in Port-au-Prince in hopes of bringing all the feuding parties together.

H U.N. officials take part in a conference called by Prime Minister Robert Malval. (does not follow, RTE1933)

CLASP centre for linguistic theory and studies in probability

The RTE platform: Brief evaluation

- Natural examples
  - However: Inclusion of controversial cases

The RTE platform: Brief evaluation

- Natural examples
  - However: Inclusion of controversial cases

P Wal-Mart is being sued by a number of its female employees who claim they were kept out of jobs in management because they were women.

H Wal-Mart is sued for sexual discrimination.

- Quite specialized and refined legal knowledge is needed for a system to infer this

CLASP centre for linguistic theory and studies in probability

The RTE platform: Brief evaluation

- Complicated in terms of their syntax and semantics given the open text nature but:
  - Simple at the same time, given that in most of the cases no "deep" (in many quotes) reasoning is involved in the examples (at least the kind discussed by formal semanticists)

centre for linguistic theory and studies in probability

The RTE platform: Brief evaluation

- This surface complexity is evident in the poor performance of all systems evaluated on the RTE tasks

CLASP centre for linguistic theory and studies in probability

The RTE platform: Brief evaluation

- This surface complexity is evident in the poor performance of all systems evaluated on the RTE tasks
- No distinction between various types of inference (logical, presuppositional, enthymematic etc. (this is to a extent a drawback of the FraCaS))

CLASP centre for linguistic theory and studies in probability

The RTE platform: Brief evaluation

- This surface complexity is evident in the poor performance of all systems evaluated on the RTE tasks
- No distinction between various types of inference (logical, presuppositional, enthymematic etc. (this is to a extent a drawback of the FraCaS))
- Sometimes unreasonable amounts of lexical knowledge are needed

centre for
linguistic theory
and studies in probability

The RTE platform: Brief evaluation

- This surface complexity is evident in the poor performance of all systems evaluated on the RTE tasks
- No distinction between various types of inference (logical, presuppositional, enthymematic etc. (this is to a extent a drawback of the FraCaS))
- Sometimes unreasonable amounts of lexical knowledge are needed
- For models that rely on big training sets, RTE is extremely small a dataset (less than 1000 cases for each suite)

CLASP centre for linguistic theory and studies in probability

The RTE platform: Brief evaluation

- This surface complexity is evident in the poor performance of all systems evaluated on the RTE tasks
- No distinction between various types of inference (logical, presuppositional, enthymematic etc. (this is to a extent a drawback of the FraCaS))
- Sometimes unreasonable amounts of lexical knowledge are needed
- For models that rely on big training sets, RTE is extremely small a dataset (less than 1000 cases for each suite)

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Developed at Stanford (Bowman et al. 2015)

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Developed at Stanford (Bowman et al. 2015)
  - Created using crowdsourcing (mechanical turk)

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Developed at Stanford (Bowman et al. 2015)
  - Created using crowdsourcing (mechanical turk)
  - The subjects are given the caption of a picture and are asked to provide:

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Developed at Stanford (Bowman et al. 2015)
  - Created using crowdsourcing (mechanical turk)
  - The subjects are given the caption of a picture and are asked to provide:
    - An alternate true caption
    - An alternate possibly true caption
    - An alternate false caption

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Developed at Stanford (Bowman et al. 2015)
  - Created using crowdsourcing (mechanical turk)
  - The subjects are given the caption of a picture and are asked to provide:
    - An alternate true caption
    - An alternate possibly true caption
    - An alternate false caption

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Instructions used on Mechanical Turk

We will show you the caption for a photo. We will not show you the photo. Using only the caption and what you know about the world:

- Write one alternate caption that is **definitely** a **true** description of the photo. *Example: For the caption "Two dogs are running through a field." you could write "There are animals outdoors."*
- Write one alternate caption that **might be** a **true** description of the photo. *Example: For the caption "Two dogs are running through a field." you could write "Some puppies are running to catch a stick."*
- Write one alternate caption that is **definitely** a **false** description of the photo. *Example: For the caption "Two dogs are running through a field." you could write "The pets are sitting on a couch." This is different from the maybe correct category because it's impossible for the dogs to be both running and sitting.*

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Huge dataset
  - Good for training NN models

The SNLI platform

- Huge dataset
  - Good for training NN models
- Inference examples are grounded in situations
- Validation: subset of the sentences were sent to a validation task

The SNLI platform

- Huge dataset
  - Good for training NN models
- Inference examples are grounded in situations
- Validation: subset of the sentences were sent to a validation task
- Misses a number of more generic inferences
  - Everything is tied to situations
  - For example: what would be the caption of *all men are human*?

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Huge dataset
  - Good for training NN models
- Inference examples are grounded in situations
- Validation: subset of the sentences were sent to a validation task
- Misses a number of more generic inferences
  - Everything is tied to situations
  - For example: what would be the caption of *all men are human*?

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Again, it seems that this platform also misses parts of the whole range of phenomena associated with NLI

centre for
linguistic theory
and studies in probability

The SNLI platform

- Again, it seems that this platform also misses parts of the whole range of phenomena associated with NLI
- On a theoretical basis, a platform that captures the complexity and coverage of NLI is not yet to be found

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Again, it seems that this platform also misses parts of the whole range of phenomena associated with NLI
- On a theoretical basis, a platform that captures the complexity and coverage of NLI is not yet to be found
  - We are currently thinking about this along with Robin Cooper and Simon Dobnik but I will say no more here

CLASP centre for linguistic theory and studies in probability

The SNLI platform

- Again, it seems that this platform also misses parts of the whole range of phenomena associated with NLI
- On a theoretical basis, a platform that captures the complexity and coverage of NLI is not yet to be found
  - We are currently thinking about this along with Robin Cooper and Simon Dobnik but I will say no more here

CLASP centre for linguistic theory and studies in probability

Formal Models of Dialogue

- Should dialogue data be part of the range of phenomena that traditional models of syntax/semantics can capture?
  - If yes, what kind of extensions to the models are needed?
    - ★ Ellipsis as a syntactic (e.g. Merchant, Kobele), semantic or more modular phenomenon (e.g. Ginzburg's work)?
    - ★ Dialogue data as a test case for formal linguistic frameworks?
    - ★ Other approaches: Poesio and Rieser (incremetal LTAG plua a model of dialogue coordination), Kempson et al. (Dynamic Syntax, Incremental model, underspecification + update as central)
  - Are these attempts of any practical relevance to dialogue systems?
  - If yes, on what level?
    - ★ A suggestion to this end: Oliver Lemon's group at Herior Watt

CLASP centre for linguistic theory and studies in probability

Formal Models of Dialogue

- Dimitrios Kalatzis, Arash Eshghi, and Oliver Lemon. Bootstrapping incremental dialogue systems: using linguistic knowledge to learn from minimal data. NIPS workshop on Learning Methods for Dialogue 2016
  - Reinforcement learning + an incremental model of syntax with a richly typed semantic backbone (DS-TTR)
- Yanchao Yu, Arash Eshghi, and Oliver Lemon. Incremental generation of visually grounded language in situated dialogue. In Proceedings of INLG 2016, Los Angeles, 2016.
  - Interactive learning of grounded word meanings. DS-TTR + visual classifiers

CLASP centre for linguistic theory and studies in probability

Grammatical Framework + semantics

- Ranta's platform for multilingual translation
  - Quite successful in multilingual translation
  - One abstract syntax, linearizations in different concrete syntaxes
  - Support for more than 20 languages
  - Recent project: from GF to UD
- But: no semantics!
- Trying to provide dependent type semantics (within the Martin Löf tradition)
- Outputing this semantics to a dependently typed proof assistant (Coq) to be reasoned about
- Accuracy very high on the FraCaS (sections 1,2,4,5)
  - In progress but a preliminary incomplete run shows accuracy above 90%
- No idea where this will lead to!

CLASP centre for linguistic theory and studies in probability