# Detecting semantic similarity between human-articulated statements and Quranic verses

Abdulrahman Alnaeem, Waad Alhoshan & Riza Batista-Navarro
School of Computer Science
University of Manchester

CLASP Seminar
University of Gothenburg
19 Dec 2018

The Quran

- book of religious teachings of Islam
- believed to be the word of God revealed to Prophet Muhammad in clear Arabic
- contains:
  - 114 chapters (Sura)
  - 6, 236 verses

Source: https://www.flickr.com/photos/el7bara/45540389/
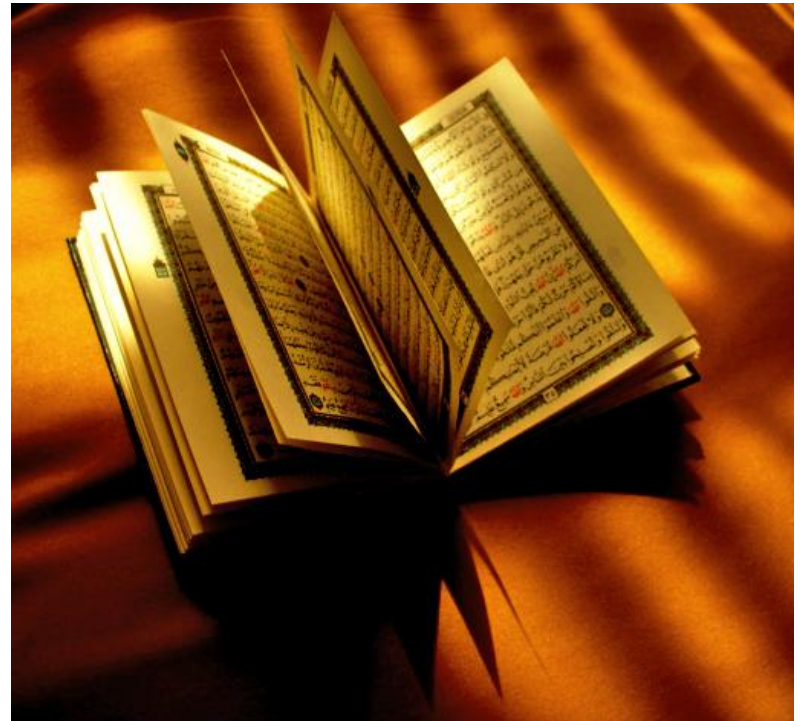
The Quran : misinterpretation

Quran 5:51

guardian, protector, advocate

> " O you who believe! do not take the Jews and the Christians for friends; they are friends of each other; and whoever amongst you takes them for a friend, then surely he is one of them; surely Allah does not guide the unjust people.

Quran 48:29

resist skepticism

non-Muslims posing as Muslims

> " O you who believe! fight those of the unbelievers who are near to you and let them find in you hardness; and know that Allah is with those who guard (against evil).

Source: http://aboutislam.net/reading-islam/understanding-islam/7-most-misinterpreted-verses-of-the-quran/

Linguistic sophistication of the Quran

- *"its style is unrivalled, its language eloquent, its meaning deep"* (Pickthall, 2004)
- *"a sea of rhetoric"* (Abdul-Raof, 2003)
- uses **metaphors** for emphasising the importance of faith, e.g., to:
  - impart wisdom
  - imply a warning
  - cite examples
  - draw images

Linguistic sophistication of the Quran

● Quranic metaphors

"...a very important cultural stock that writers and public speakers keep falling back on in cases where they need to prove a point or establish an indisputable point of view. Speakers quote these (metaphorical) verses to convince the addressee that what they say is guaranteed to be true."

(Zahir, 1991)

Metaphor

- figure of speech comparing two dissimilar items that have something in common
- the *vehicle* carries a number of components of meaning of which usually only one is contextually relevant to and shared by the *topic* (Beekman and Callow, 1986)
- *grounds* (**context**) help in recognising intended meaning (Knowles and Moon, 2006)

> topic: "life"
> vehicle: "box of chocolates"
> grounds: "you never know what you are going to get"

Metaphor

- Conceptual Metaphor Theory (CMT)

  "... understanding meaning from one domain (that is typically abstract) in terms of another (that is typically concrete)."

  "...mapping from sensorimotor domains to domains of subjective experience"

Aims and objectives

- to detect misinterpretation of Quranic verses
- to determine how close a human-articulated statement is to its cited Quranic verse
- to develop an approach based on the measurement of similarity between any given pair of texts:
  - a human-articulated statement citing a Quranic verse
  - a verse from what is considered as a gold standard English version of the Quran

Semantic Textual Similarity (STS)

Compute the similarity between two pieces of text S1 and S2

Shared Tasks (2012-2017)

- state of the art (SotA)
    - deep learning models using sentence embeddings
    - 85% Pearson correlation on English pairs
- Interpretable STS
    - explains similarities/differences
    - identification and alignment of chunks
    - assignment of relation and similarity score to each alignment
    - SotA: 64% (using distributional similarity models)

Interpretable STS

S1:

"12 killed in bus accident in Pakistan"

S2:

"10 killed in road accident in NW Pakistan"

[12] **Similar (4)** [10]

[killed] **Equivalent (5)** [killed]

[in bus accident] **More specific (4)** [in road accident]

[in Pakistan] **More general (4)** [in NW Pakistan]

Proposed approach

Development of an annotated corpus

Interpretable STS

- detecting and interpreting metaphors
- using embedding-based representations

Annotated corpus

Sources

- gold standard English version of the Quran: King Fahd Complex for the Printing of the Holy Quran (Medina, Saudi Arabia)
- websites, blog posts, news articles

Pairs of verses

Indeed My lord is the hearer of supplication (14:39)
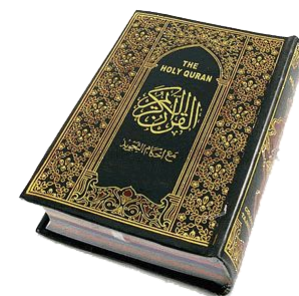
And they have attributed to Allah equals to mislead (people) from His way. Say, 'Enjoy yourselves, for indeed, your destination is **the Fire**.' (14:39)

Annotated corpus

Pairs of verses annotated as:
- lexically equivalent
- semantically equivalent
  - how?
- not equivalent

Indeed My lord is the hearer of supplication (14:39)

And they have attributed to Allah equals to mislead [people] from His way. Say, 'Enjoy yourselves, for indeed, your destination is **the Fire**.' (14:39)

Metaphor detection: Recent work

2018 VUA Metaphor Detection Shared Task

- ~65% F-score (all POS), ~67% F-score (verbs)
- pre-trained word embeddings (word2vec), bidirectional long short-term memory (Bi-LSTM)

Adjective-Noun (AN) pairs (Bizzoni et al, 2017)

- 91% accuracy
- pre-trained word embeddings (word2vec)

Joint metaphor detection & interpretation

Mao et al, 2018

- applied to English-Chinese machine translation (MT)
- word embeddings trained on Wikipedia (word2vec)
- WordNet lookup
- 62% (Google) and 65% (Bing) accuracy

Proposed approach: Frame embeddings

Semantic frames

- allows for analysis of the internal meaning of a metaphor, by taking their components into account (Gemmel, 2015)
- using FrameNet
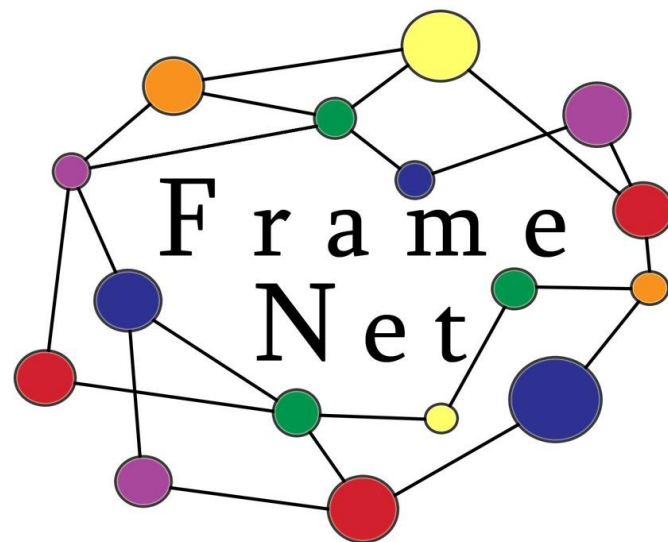
Frame embeddings: Alhoshan, 2018

Requirements Engineering (RE)

● concerned with discovering, developing, tracing, analysing, and managing requirements that define a system at successive levels of abstraction
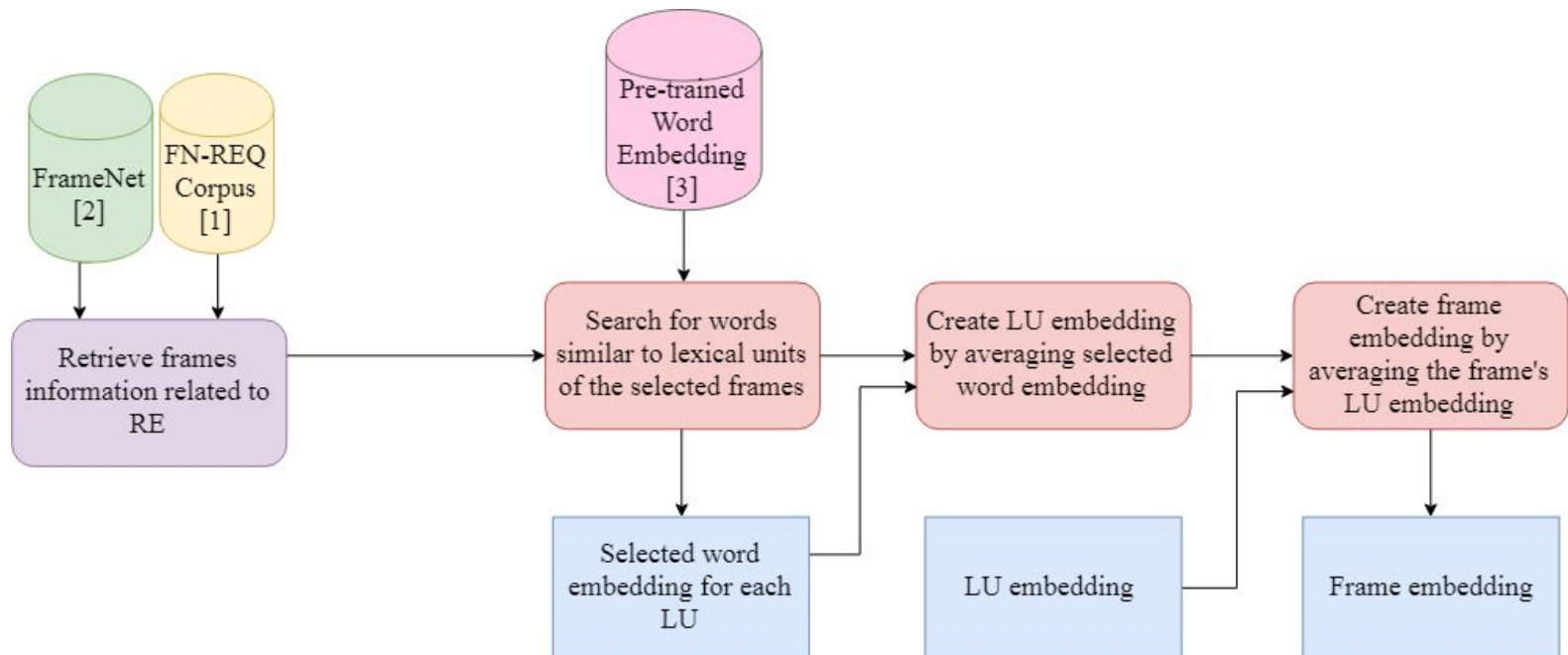
Frame embeddings: Alhoshan, 2018

FN-REQ Corpus

Annotation Level	IAA (F-score)
Evoked frames	78.50%
Frame elements (roles only)	76.90%
Frame elements (roles and text spans)	72.85%

Frame embeddings: Alhoshan, 2018

Requirements Engineering (RE)

- pre-trained word2vec model on 1.0 GB of Stack overflow technical posts
- applied to a task for measuring frame semantic relatedness
- ongoing work: comparison with results from training embeddings on real RE documents

Frame embeddings: Ongoing work

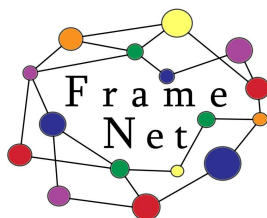*Say, 'Enjoy yourselves, for indeed, your destination is the <u>Fire</u>.' (14:39)*

Training two sets of frame embeddings

- based on frames identified over non-metaphors
- based on frames identified over metaphors

Frame embeddings: Ongoing work

Say, 'Enjoy yourselves, for indeed, your destination is the <u>Fire</u>.' (14:39)

Non-metaphorical language

Metaphorical language

Fire_burning *(Fire[Act], <u>Fuel</u>)*

We watched in horror as [tanker] [fire] <u>roasted</u> female driver...

Four people have <u>died</u> so far in the [<u>blazes</u>].

Yesterday, [<u>flames</u>] continued to threaten Galicia.

We will drive them into a [Fire]. Every time their skins are <u>roasted</u> through...(4:56)

Surely, God has cursed the disbelievers, and has prepared for them a <u>flaming</u> [Fire] wherein they will abide forever. (33:64)

Jahīm: A <u>blazing</u> [fire].

Open questions and conclusion

- Focus on specific types of metaphors only?
- Available metaphor-annotated corpora?
- Potential collaborations?

Thank you!