Similarity (Metric) Learning

Stefan Mojsilovic[1]

[1]R&D Centre - Belgrade
Everseen Ltd.

MATF, Belgrade, May 2021.

Outline

Who am I?

- Bio:
  - BSc in Signals and Systems (Automation) from ETF, Serbia
  - MSc in ML from Aalto University, Finland.
  - Teaching assistant at Aalto University, Finland.
  - Research assistant at University of Helsinki, Finland.
  - ML Engineer, AI researcher, Tech Lead, Team Leader at Everseen for about 2.5 years.
  - Lecturer and mentor at PSIML 2020. and 2021.
- AI Interests:
  - Computer Vision, Representation Learning;
  - Augmented Intelligence - how to use AI to extend our own?
  - Causal Learning, Evolutionary Algos, Reinforcement; Learning;
  - Neuroscience, Psychology, Philosophy.

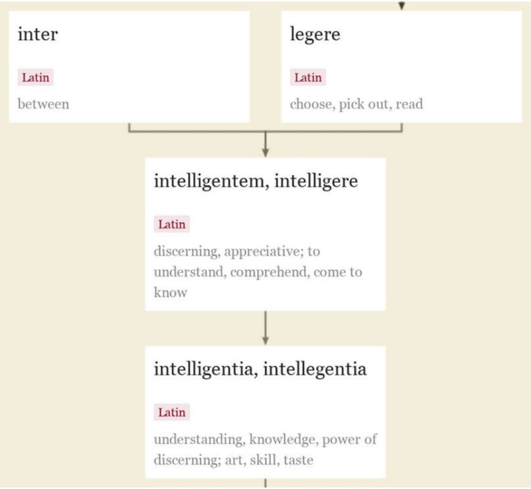
Everseen

- Applications of ML and CV in the retail industry
- Our products are deployed in 1000s of stores on 4 continents
- Belgrade R&D
  - Product Switch
  - Non-Scan
  - Basket/Cart based loss
  - Transaction analysis
- more at `https://everseen.com/`

Motivation

- Goal: Learn the notion of similarity in computer-based systems
- Why?

Etymology


inter
Latin
between
legere
Latin
choose, pick out, read
intelligentem, intelligere
Latin
discerning, appreciative; to understand, comprehend, come to know
intelligentia, intellegentia
Latin
understanding, knowledge, power of discerning; art, skill, taste

What is similarity?

- Similar (adj.) - Having a resemblance in appearance, character, or quantity, without being identical - "Having characteristics in common".
- "Similarity between objects plays an important role in both human cognitive processes and artificial systems for recognition and categorization." Bellet et al. [2015]

Applications

- Information Retrieval / Search engines
  - text, images, music...
- Recommender systems
  - products, content, services, people...
- Verification / Re-identification
  - people, cars, objects...
- Unsupervised ML algorithms and Nearest Neighbor methods

How do we measure similarity? (1/2)

- Objects represented as sets of characteristics (features).
- Similarity of objects as real-valued bivariate functions defined on pairs of such sets (using norm, intersection, difference, etc.).
  - Jaccard index
  - SorensenDice coefficient
  - Overlap coefficient
  - Tversky index (a generalization of the SorensenDice coefficient and the Tanimoto coefficient (aka Jaccard index))

How do we measure similarity? (2/2)

- You want to buy a pet - a Sphynx cat. However, in the pet store you have to choose between:
  - a stuffed Sphynx cat looking exactly as you would like it,
  - a Chihuahua looking very similar to the sphinx cat,
  - a very different kind of a cat, say a Maine Coon.
- How would you choose the most similar pet?
- How would you quantify those similarities on a $[0, 1]$ interval?




Formulation (1/2)

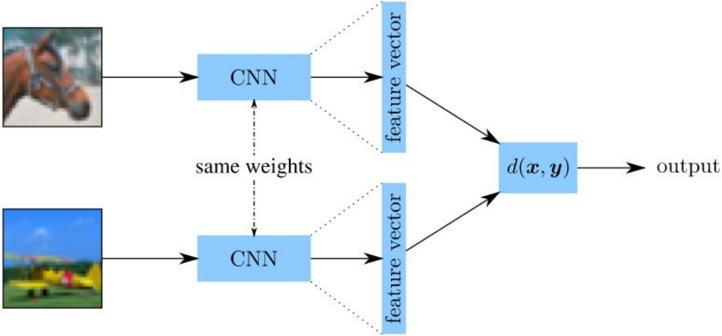
- Goal: Learn the notion of similarity in computer-based systems.
- Qualifying similarity via a mapping from pairs of inputs to $\{\text{similar}, \text{dissimilar}\}$ or
- Quantifying via mapping to e.g. $[-1, 1]$ or $[0, 1]$ - higher for more similar and lower for less similar.
- Proposition: Parametrize the mapping as a neural network and learn the parameters to optimize for the desired outcome.
  - Map objects to an embedding (feature) space $\mathbb{E} \subseteq \mathbb{R}^n$ and use predefined measures (e.g. Euclidean distance, Cosine similarity);
  - Learn the similarity/distance measure on top of such embeddings.
- Problem: How to obtain labels?

Formulation (2/2)

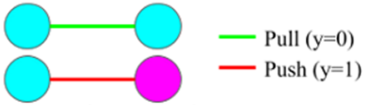
- Absolute similarity assessment is very difficult and unreliable for humans - "A and B are 0.83 similar".
- Relative similarity assessment comes naturally:
  - "A and B are similar. C and D are dissimilar." - Context?
  - "A and B are more similar than A and C". A serves as a contextual anchor.

Siamese Networks

Bromley et al. [1993] Taigman et al. [2014]



Contrastive Loss (1/2)



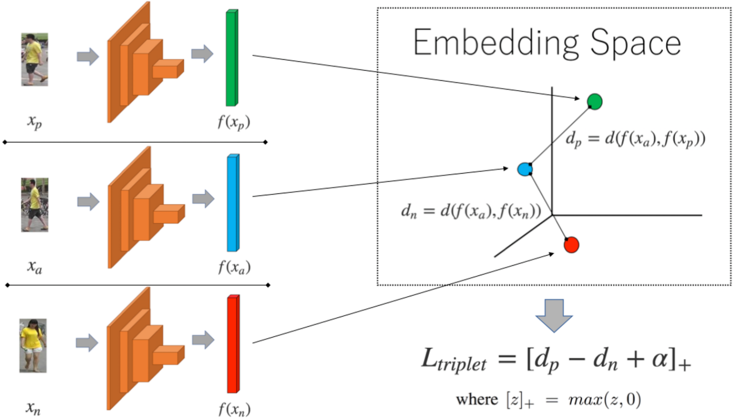
$$L_{\text{contrastive}} = \underline{(1-y) \times D(x_i, x_j)} + \underline{y \times \max(0, m - D(x_i, x_j))}$$

Contrastive Loss (2/2)

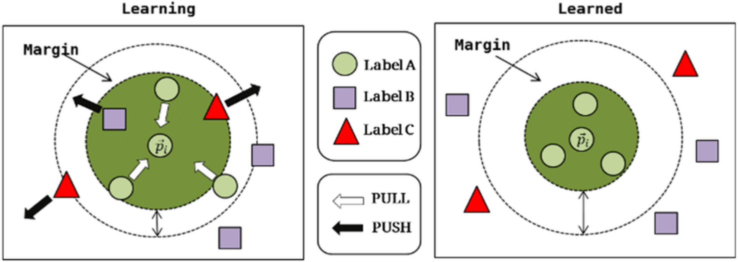
- Pairwise Loss - N^2 pairs
- Context cannot be inferred from the pair at hand
- Wants to collapse objects belonging to same-class pairs

Triplet Networks and Loss (1/3)

Chechik et al. [2010] Schroff et al. [2015]

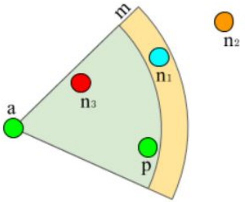


Triplet Networks and Loss (2/3)



Triplet Networks and Loss (3/3)

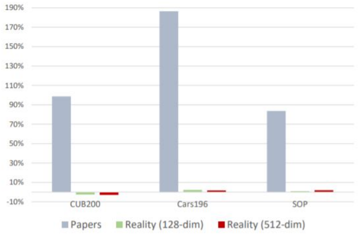
- Pro: More fine grained (N^3 vs N^2)
- Pro: Context provided via the anchor object
- Con: Depends heavily on triplet mining strategies



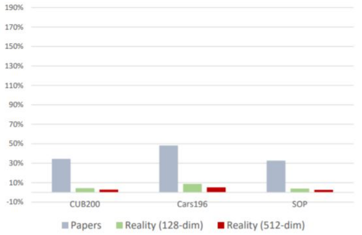
Triplet loss tuple (anchor, positive, negative) and margin m. Hard, semi-hard and easy negatives highlighted in red, cyan and orange respectively.

Literature Survey

- Survey - Kulis et al. [2012]
- Book - Bellet et al. [2015]
- Reality check - Musgrave et al. [2020]
- Github - pytorch-metric-learning



(a) Relative improvement over the contrastive loss



(b) Relative improvement over the triplet loss

Summary

- Assessing similarities is an important part of AI.
- Widely used in real-world applications.
- Similarity depends on context and relative similarities come more naturally.
- Siamese and Triplet Networks as SOTA approaches.
- The field is still growing and contributions are welcome.

References I

Aurélien Bellet, Amaury Habrard, and Marc Sebban. *Metric Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2015. doi: 10.2200/S00626ED1V01Y201501AIM030. URL https://doi.org/10.2200/S00626ED1V01Y201501AIM030.

Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, 6:737–744, 1993.

Gal Chechik, Varun Sharma, Uri Shalit, and Samy Bengio. Large scale online learning of image similarity through ranking. *J. Mach. Learn. Res.*, 11:11091135, March 2010. ISSN 1532-4435.

References II

Brian Kulis et al. Metric learning: A survey. *Foundations and trends in machine learning*, 5(4):287–364, 2012.

Kevin Musgrave, Serge Belongie, and Ser-Nam Lim. A metric learning reality check, 2020.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015. URL http://arxiv.org/abs/1503.03832.

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014. doi: 10.1109/CVPR.2014.220.