UNIWERSYTET EKONOMICZNY W KRAKOWIE

Towards human perception of relatedness: a fuzzy semantic relatedness measure

Grażyna Paliwoda-Pękosz

Cracow University of Economics, Poland

Agenda

- Semantic relatedness
- Semantic relatedness measures
- Fuzzy sets
- Linguistic variables
- Proposition of a fuzzy semantic relatedness measure
- Areas of applications and future research

Semantic Relatedness

- Semantic similarity stems from lexical relations:
  - synonymy (e.g. car, auto)
  - hyponymy (“is a”)/hypernymy (“more general than”) e.g. house, building
- Semantic relatedness – a more general term than semantic similarity, encompasses:
  - meronymy (“part of”, e.g. wheel, car)
  - antonymy (e.g. hot, cold)
  - associations (e.g. student, university)
- Semantic distance – the inverse of semantic relatedness
- Semantic relatedness application: information retrieval, word-sense disambiguation, ontology matching, ontology based data analysis

Semantic Relatedness Measure

$$rel: O \times O \rightarrow [0, 1]$$

where: O – a set of words, documents, graphs, …

rel(x,y)=1 if and only if x=y

- Kinds of SRMs:
  - vector-based measures (e.g. Latent Semantic Analysis)
  - graph-based measures (Rada, Leacock & Chodorow, Hirst&St-Onge, Wu&Palmer)
  - information theory – based measures (Resnik, Lin)
  - gloss-based measures (Lesk)
  - hybrid approaches
- Language resources
  - Wordnet, Wikipedia, WWW, …

Fuzzy sets

- Set of companies that are quoted on the Warsaw Stock Exchange
  => a crisp set
- Set of companies that have high earnings
  => ?

$$A = \{(x, \mu_A(x)): x \in X\}$$

where:

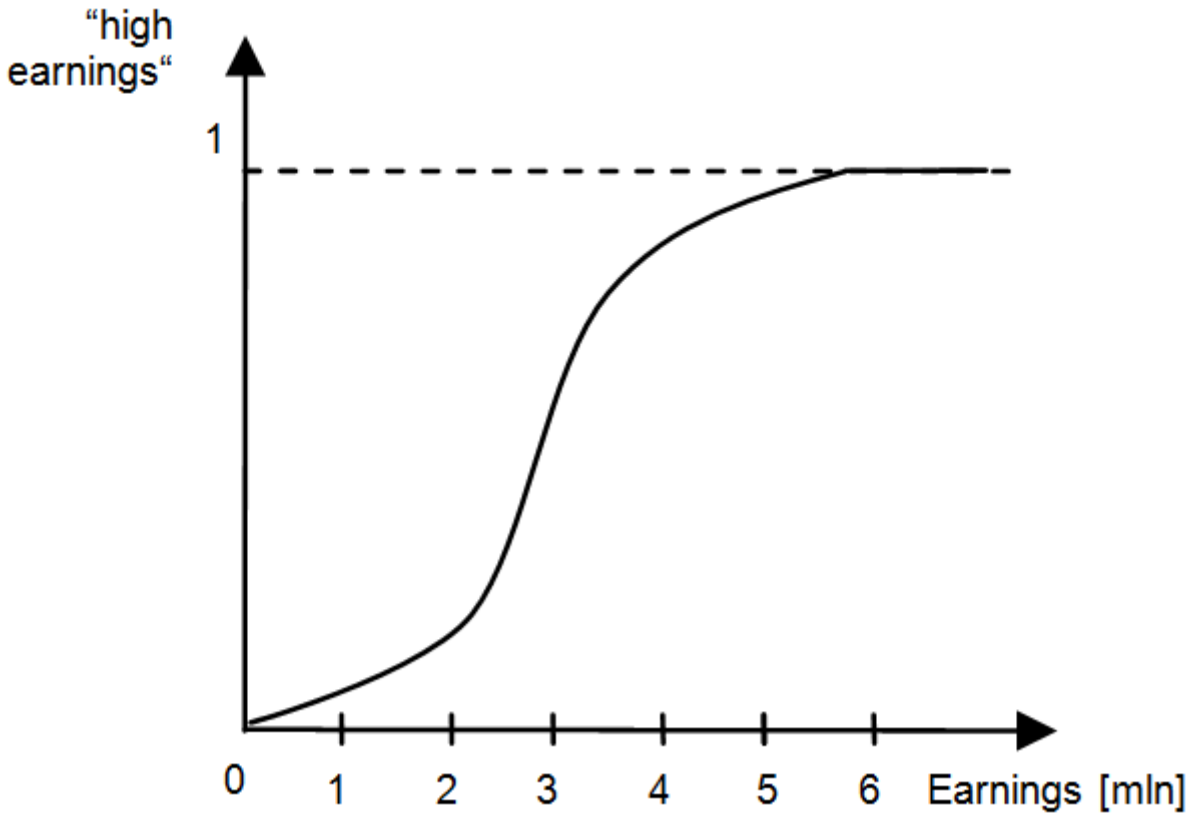
A – a fuzzy set defined in a universe of discourse X

μ_A – a membership function, $\mu_A: X \rightarrow [0,1]$

Zadeh, L. (1965). Fuzzy sets. Information and Control (8), 338-353.

Fuzzy sets example

- An example of a membership function for „high earnings“



Linguistic variable

$$LV = (N, T(N), U, S, M)$$

where:

LV – a linguistic variable

N – a name of a variable

T(N) – a collection of possible linguistic values for N

U – a universe of discourse

S – a syntactic rule that generate terms in T(N)

M – a semantic rule that for each value of a linguistic variable x assigns its meaning M(x) that is defined by a fuzzy subset of U. This fuzzy subset is characterised by a compatibility membership function: $c:U \rightarrow [0,1]$

Zadeh, L. (1975). The concept of a linguistic variable and its applications to approximate reasoning, Part I. Information Sciences (8), 199–249.

Proposition of a fuzzy semantic relatedness measure

(FSRM, T(FSRM), U, S, M)

where:

FSRM – a fuzzy semantic relationships measure

T(FSRM) – a collection of possible values of a fuzzy semantic relationships measure

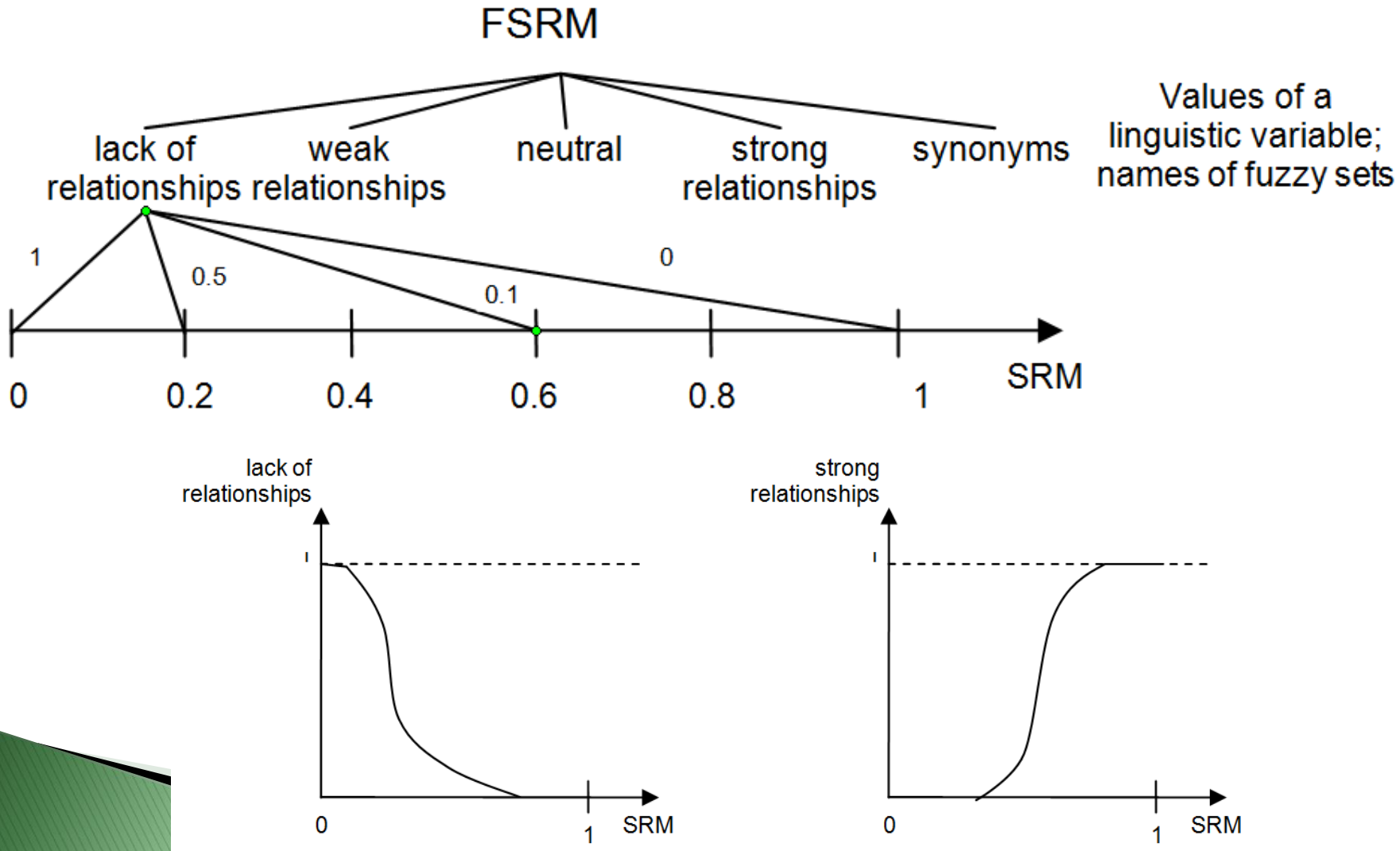
T(FSRM)= lack of relationships + weak relationships + neutral + strong relationships + synonyms

U – a universe of discourse; a set of words or documents for which semantic relatedness will be assessed

S – a syntactic rule

M – a semantic rule; M(x,y)– a meaning of (x,y) (a fuzzy subset of U) – a fuzzy set that is defined by a compatibility function which assess a semantic relatedness measure (SRM) between x and y:
$\forall$ (x,y) $\in$U M(x,y)=SRM(x,y)

A fuzzy semantic relationship measure as a linguistic variable



Areas of applications and future research

- Systems that have to simulate human perception of semantic relatedness
- Automatisation of the process of human reasoning
- Future research: Information filtering model based on the concept of fuzzy semantic relatedness measure

UNIWERSYTET EKONOMICZNY W KRAKOWIE

Thank you

Towards human perception of relatedness: a fuzzy semantic relatedness measure

Grażyna Paliwoda-Pękosz

Cracow University of Economics, Poland