Approximating Life Insurance Liabilities – Viewed from a Machine Learning Angle

SAV – Arbeitsgruppentagung AFIR

Lugano

September 1, 2017

Stefan Jaschke

infinada

Outline

1. Machine Learning / Artificial Intelligence

- top-down business-view
- bottom-up technical view

2. Approximating Life Insurance Liabilities

- Solvency 2 (Europe), AG43 & C3P2 (US)
- ALM for life insurance landscape: Moody's, WTW, Milliman, Deloitte
- academic landscape
- perceived trends
- ideas and theses

Machine Learning (Business View)

Still small, but growing fast*:

- $26bn – $39bn invested in 2016
- 1-3% of total tech investment
- annual growth rate of capital invested by Venture Capital firms is about 35 – 40% in the period 2010-2016

Spending is dominated by the “tech giants” (like Google, Amazon, Apple, Baidu, Facebook, IBM, Microsoft, Netflix, ...):

- $20bn – $30bn invested in 2016 by the 35 most important firms

Adoption of Artificial Intelligence (AI) in three tiers:

(1) tech & telecoms, automotive, financial services

(2) retail, media & entertainment

(3) education, health care, travel

Many companies are still monitoring and experimenting.

**McKinsey Global Institute: “Artificial intelligence – The next digital frontier”, June 6, 2017, and “Smartening up with artificial intelligence – What’s in it for Germany and its industrial sector”, April 19, 2017*

Narrow AI/ML Application Breakthroughs

Breakthroughs:

- image recognition, computer vision
- natural language processing
- knowledge representation and fast retrieval

Improvements:

- remarkable improvements in deriving predictions from large data bases

Highly Visible Applications

- autonomous vehicles (cars, copters, drones)
- Siri, Cortana, Google Now!
- IBM Watson
- AlphaGo (2016), Libratus (no-limit Texas Hold'em, 2017)

Less widely publicized:

- classification in insurance: automated underwriting, claims management, fraud detection
- classification in health care: X-ray and MRI images, gene expression data, ...

Machine Learning (Bottom Up)

“System II” (Kahneman):

- analytical, using words and language
- can reason with logic
- in humans, not animals:
- causation = understand a statistical result, build theories

“System I”:

- fast, creative, associative, passionate
- can reason with intuition
- done well by humans and animals:
- correlation = uncover a statistical result, recognize patterns

Cognitive Computing:

- knowledge graphs, ontologies
- automatic reasoners
- shallow reasoning can be parallelized efficiently

Statistical Learning:

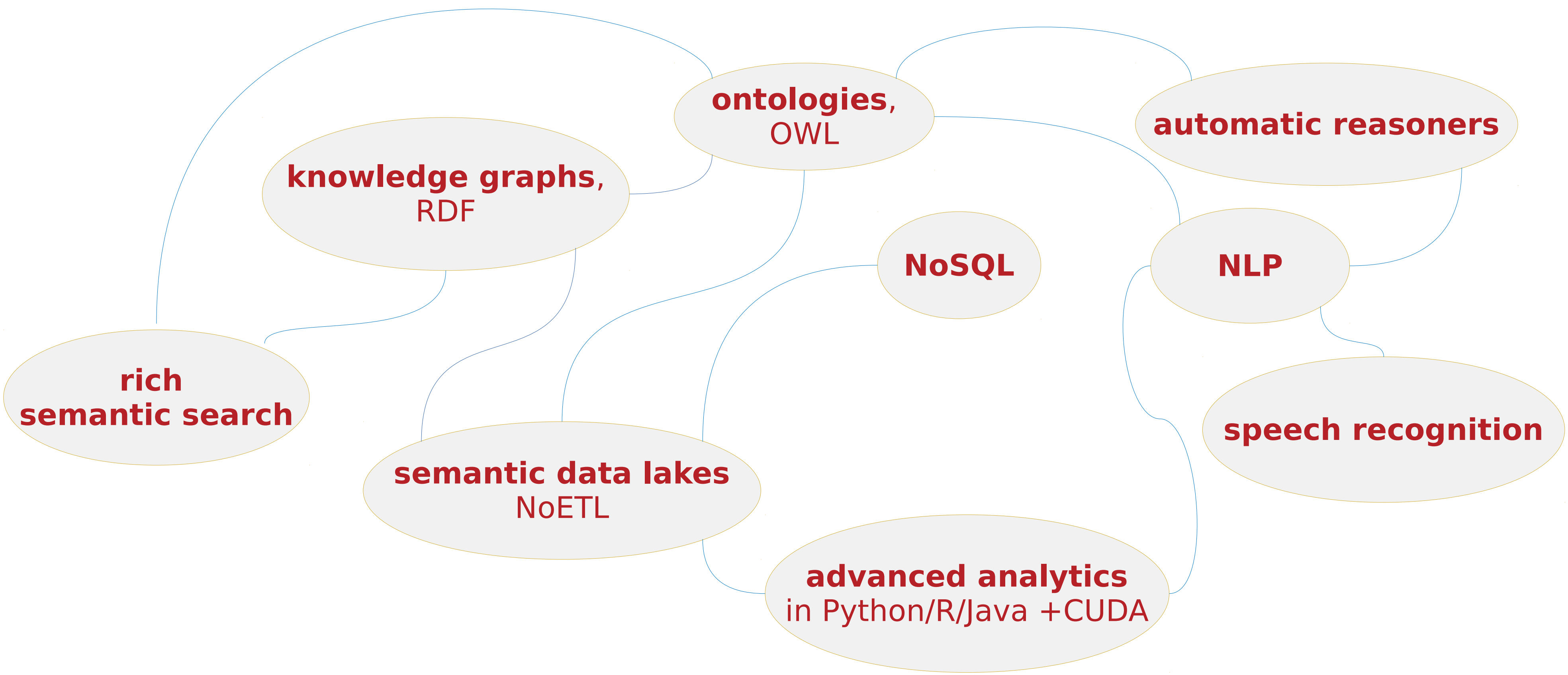
- deep neural nets, SVMs, random forests, … nonlinear methods
- massively parallel computations: GPUs, FPGAs, TPUs, “neuro”-chips

Cognitive Computing: Concepts & Keywords

Beneath the **tip of the iceberg**:

- artificial intelligence & machine learning: Apple's Siri, IBM's Watson, Google's semantic search, …

is the less visible “**cognitive computing**”:



What is an Ontology?

How do different tribes *use* "ontology"?

Three shades of meaning (partly overlapping):

1. Provide "meaning" in the knowledge graph. This is the basis for "rich semantic search" and this is what Google's knowledge graph supports: E.g. Google knows two meanings of "Rosetta Stone": the artifact and the software.

http://google.com/?q=Rosetta+Stone

2. Agree on a common vocabulary and consistent rules for applying this vocabulary. Examples in medicine: "ICD-10" or "SNOMED" or the financial industry: "FIBO". The agreed upon common vocabulary acts as the "Rosetta Stone" to connect different sources of knowledge.

3. Define rules to enable automatic reasoning. Use languages that are expressive enough to be interesting but limited enough to allow practicable "theorem provers" (= automatic reasoners).

Specific Knowledge Graphs (2015-2017)

	facts	entities	remark
Google KG	70bn (2016)	1bn (2015)	uses “component value types” to encode n-ary relations for n>2
Metaweb Freebase	1.9bn (2015)	48m (2015)	acquired by Google 2010; phased out Jun 2015 (and small portion transferred to Wikidata)
Wikimedia Foundation Wikidata	600m triples (2015)	14m (2015)	entities can act as classes; predicates can have further “qualifiers”; facts are independent of the language (covers 400 languages); hosted by Blazegraph
FU Berlin + Univ. Leipzig DBPedia	580m (engl.)	4.6m	“knowledge” extracted from Wikipedia; unlike Wikidata, additional triples for different languages might be inconsistent; hosted by Virtuoso
mpii Yago3	120m	10m	“knowledge” extracted from Wikipedia, wordnet and GeoNames; hosted by Virtuoso
MusicBrainz	?	21m	facts about 17.7m recordings, 1.9m releases and 1.2m artists; originally hosted in PostgreSQL
Cyc	15m	700k	Cyc started 1984 as an attempt to build an ontology of everyday common sense knowledge
Wolfram Alpha	?	?	question-answering machine, very likely employing some form of knowledge graph

Cognitive Computing Business Models

firms: products	main story lines
Apple: Siri, Microsoft: Cortana, Google Now!	digital personal assistants with a natural language interface; often related to search
IPSoft: Amelia, inbenta: Veronica, ExpectLabs (now Cisco): MindMeld	help-desk and phone center support, customer service apps; adding voice support to mobile apps
IBM: Watson, Palantir: Gotham, Saffron (now Intel): Saffron cognitive solutions (associative memory)	flag patterns in large knowledge bases: policing & anti-terror intelligence, fraud detection, customer relationship mgmt. (CRM), software life-cycle support, ...

Selected Business Models*

firms: products	main story lines
DataStax: Cassandra	transactional databases with master-master replication => full **P**artition tolerance and high **A**vailability; fully "ACID" with eventual and tunable **C**onsistency
declara, InsightNG	personalized learning and knowledge discovery
CoherentKnowledge: ERGO (Rulelog)	"understanding" and reasoning over legal language, e.g. financial regulatory texts and transactions
Capsicum (now Capsifi): Jalapeno	business architecture consulting: FIBO, ISO20022
Capsenta: Ultrawrap	"NoETL" federation of different data sources

** SmartData Week, San Jose, August 2015*
SmartData Conference, Redwood City, January 2017

Approximation of Life Insurance Portfolios

Use cases:

- SCR in internal models in Solvency II
- reserving (AG43) and capital requirements (C-3 Phase2) for US variable annuities
- liability driven investment, ALM
- hedging platform: hundreds of greeks for dozens of portfolios
- pricing of reinsurance contracts (base price + risk margin)

Usual classification:

practice: hybrid of all 3!

"pricing surface" story

	outer scen.	inner scen.	function base, feature selection	
curve fitting	few (~200)	many (~5k)	simple, analytic functions (e.g. quadratic)	trading & hedging operations, pricing
replicating portfolio	more (5k..50k)	4..50	"meaningful" analytic functions, "features" pre-selected by humans	typical software: Algorithmics or in-house
LSMC	many (10k..2m)	few (1..4)	some function base, machine-selected "features"	Milliman, Moody's, WTW, Deloitte, ...

Approximate PV-Surfaces with LSMC

[1] S. Morrison: “Nested Simulation for Economic Capital”, Barrie+Hibbert Insights, Dec 2009.

[2] A. Koursaris: “The Advantages of LSMC”, B+H Insights, Jul 2011; “A LSMC Approach to Liability Proxy Modelling and Capital Calculation”, B+H Insights, Sep 2011.

[3] M.Hörig, M.Leitschkis: “Solvency II Proxy Modelling via LSMC”, Milliman, Jan 2012.

[4] T.Kalberer: “Stochastic determination of the VaR for a portfolio of assets and liabilities I”, Der Aktuar Q1 2012; “… – Estimation Error, II”, Der Aktuar Q2 2012; “… III”, Der Aktuar Q3 2012.

[5] M.Leitschkis, M.Hörig, F.Ketterer, C.Bettels: “LSMC for fast and robust capital projections”, Milliman Feb 2013.

[6] S.Morrison, L.Tadrowski, C.Turnbull: “1-year projection of run-off CTE reserves”, Moody’s Mar 2013.

[7] S.Morrison, C.Turnbull, N.Vysniauskas: “Multi-year projections of market-consistent liability valuations”, Moody’s Apr 2013.

[8] G. Conn: “Proxy Function Fitting: Some Implementation Topics”, Moody’s, Oct 2013.

[9] M.Hörig, K.Murray, E.Phelan, Leitschkis: “An application of Monte Carlo proxy techniques to variable annuity business: A case study”, Milliman Nov 2013.

[10] S.Morrison, L.Tadrowski: “Efficient Statistical Estimation of 1-year VaR Economic Capital”, Moody’s Nov 2013.

[11] A.Clayton, S.Morrison, C.Turnbull, N.Vysniauskas: “Proxy functions for the projection of Variable Annuity Greeks”, Moody’s Nov 2013.

[12] D.McLean, D.Redfern, K.Pyper: “A Survey of Regression Methods for Proxy Functions”, Moody’s Mar 2014. (The paper touches on GAM, neural networks, regression trees, kernel smoothing, lowess and finite elements as potential alternative to polynomials.)

[13] M.Elliot: “Making Proxy Functions Work in Practice”, Moody’s Feb 2016.

Approximate PV-Surfaces with LSMC

Main lessons:

Cashflow approximation with “replicating portfolios” may be sub-optimal for a number of reasons.

LSMC approximations of PV-surfaces are very versatile. They can be applied not only to 1-year VaR (Solv2), but also to

- CTE-based reserving (AG43) and capital calculation (C3P2),
- multi-year projections,
- the computation of coverage ratios, hedge ratios and other risk parameters.

LSMC can be fully automated (including “feature selection”) and it may be improved by

i. using very few inner scenarios per outer scenario,
ii. potentially increasing numbers of inner scenarios (dynamically) for some outer scenarios,
iii. using uniformly distributed (instead of real-world) outer scenarios,
iv. correcting the bias in the VaR estimate induced by the additional variance caused by the inner scenarios,
v. using orthogonal function bases (or bases with low condition number),
vi. using different (independent) inner scenarios for different model points.

Academic Insights

[1] M.Gordy, S.Juneja: "Nested Simulation in Portfolio Risk Measurement", FED Discussion Paper 2008-21, Apr 2008:

- use independent inner scenarios and let the number of outer scenarios grow proportionally with N^2 (N = the number of inner scenarios)
- use re-sampling to estimate bias and variance and correct for the bias; potentially use it to allocate the number of inner scenarios dynamically

[2] M.Broadie, Y.Du, C.Moallemi: "Efficient Risk Estimation via Nested Sequential Simulation", Columbia University Working Paper, 2010-12-13:

- sequential allocation of computational effort in inner scenarios
- => asymptotic error rate goes from $k^{-2/3}$ to $k^{-4/5+\varepsilon}$

[3] E.Beutner, A.Pelsser, J.Schweizer: "Fast Convergence of Regress-Later Estimates in LSMC", Maastricht University, 2013-09-20:

- **"regress-now" is fundamentally different from "regress later"**
- "regress-now" = regress a later value function or discounted cash flow onto a set of basis functions at an earlier time, e.g. t=1
- "regress-later" = fitting cash flows with explicitly known conditional expectation, for example using a replicating portfolio of assets that can be priced analytically
- "regress-later" can converge faster than 1/N
- the "regress-now" error depends on the *projection error* $||X – E_t[X]||_2$, where X is the claim and $E_t[X]$ is its conditional expectation, whereas the *projection error* is zero in "regress-later"

Academic Insights

[4] J.Natolski, R.Werner: "Mathematical analysis of different approaches for replicating portfolios", Working Paper 2014-04-16, appeared to EAJ.:

- compares the approximation of "terminal values" (like in Oechslin et.al. 2007) with the approximation of "present values"
- => If dynamic trading in the numeraire is included in the replicating portfolio, then matching terminal values is almost the same as matching present values, subject to a measure change.

[5] A.Pelsser, J.Schweizer: "The Difference between LSMC and Replicating Portfolio in Insurance Liability Modeling", Maastricht University, 2015-01-29:

- didactic introduction to the difference of "regress-now" versus "regress-later"
- main message as before: "regress-later" is potentially more powerful
- **But**: "LSMC" approaches are used slightly differently in practice (e.g. on uniform scenarios instead of on real-world scenarios) and
- "RP" approaches may not always involve analytic pricing of conditional expectations

[6] J.Natolski, R.Werner: "Replicating portfolios – interplay between objective function and numeraire", Universität Augsburg, 2015-08-15:

- extreme values in interest rates may dominate the solution => re-scale cash-flows properly by working with numeraire-discounted values
- discussion on using L^1 versus L^2 -penalties in the regression

Academic Insights

[7] E.Beutner, A.Pelsser, J.Schweizer: "Theory and Validation of Replicating Portfolios in Insurance Risk Management", Maastricht University, 2016-03-20:

- summary of previous insights
- specific examples, including a path-dependent one => avoid tensor-product-style bases for path-dependent options

[8] J.Natolski, R.Werner: "Mathematical foundation of the replicating portfolio approach", Maastricht University, 2016-04-26:

- exact error bounds for MCEV and economic capital
- using different measures for [0,1] and [1,∞]

[9] M.Cambou, D.Filipovic: "Replicating Portfolio Approach to Capital Calculation", EPFL Lausanne, 2016-05-10:

- a dynamic-hedging RP approach related to the chaos expansion of martingales on Wiener space

Open Question:

How to achieve the apparent advantages of "regress-later" in the real-world, especially in the presence of non-trivial path-dependency and if the re-calibration of the risk-neutral ESG is not directly accessible.

From nonparametric statistics to ML

step	nonparametric statistics 1997	"ML-style" statistical learning 2017
clean data	partly automated ETL into human-generated schemata	mostly automated, flagging "rare issues" for humans to look at
select meta-approach	expert human choice	attempts at automation, but still mostly human choice
select "features"	expert human choice	fully automated
fit model	automated	fully automated
cross-validate & refit	mostly automated	fully automated
validate model (narrow sense)	human analysis => feedback loop to feature selection	fully automated
compare meta-approaches	expert knowledge (fairly static, mostly asymptotic results) => feedback loop to meta-approach selection	meta-model analysis by humans => feedback loop to meta-approach selection

compare:
P ?= NP

Trends & Theses

(1) **More and more steps in the statistical learning loop become automated**. The trend towards LSMC (away from curve fitting and RP) will continue in the sense that “feature selection” and “model choice” are more and more automated and function bases and model parameters do not need to have any “meaning” for humans.

(2) **We will see other ML-inspired approaches to the approximation of life insurance liabilities** than (orthogonal) polynomials.

(3) GPU-enabled computing allows to perform computations on €10k-machines today, for which insurers have been using 100+-CPU clusters.

Once insurers switch to using GPUs, we will see different methods, software architectures and deployment processes.

New Challenges for Humans

How to trust the new "machine-learned" solutions?

What to trust them for?

What are their limitations and weaknesses?

How to find weaknesses in "machine-learned" solutions?

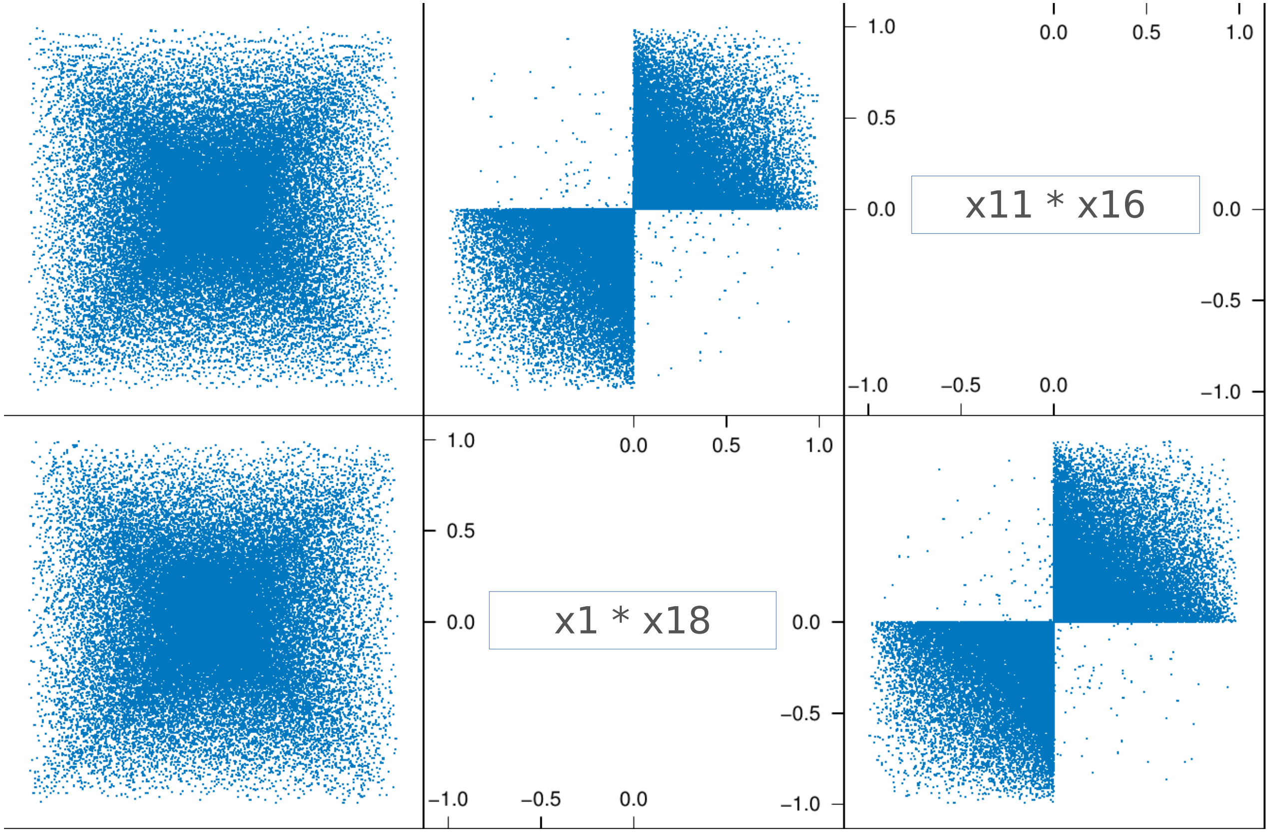
A Tesla Model S involved in the fatal crash on May 7, 2016 is shown with the top third of the car sheared off by the impact of the collision of the Tesla with a tractor-trailer truck on nearby highway and came to rest in the yard of Robert and Chrissy VanKavelaar in Williston, Florida, U.S. on May 7, 2016. Courtesy Robert VanKavelaar/Handout via REUTERS

More fully automated, "black-box"-fitting of ever more complex models requires even more thoughtful risk management and validation.

Validation Ideas

- **Use the "ML-spirit" and technology (GPUs) to find weaknesses** in machine-learned solutions.
- **Check against model criteria that are meaningful for humans**.
- **Compute bounds for key criteria** and use machine-learning technology to be able to do this.

Example:
Deficiency in Sobol numbers used as calibration scenarios:



Dr. Stefan Jaschke

- Diploma in Mathematics 1993, Humboldt University Berlin
- Fulbright Grant 1993-1994, Columbia University New York
- Dissertation 1997, Humboldt University
- 1999 – 2003 Weierstrass-Institute, Berlin: *financial mathematics*
- 2003 – 2007 BaFin, Bonn: *banking and insurance supervision*
- 2007 – 2015 MunichRe, Head of Department "Quantitative Models" in life reinsurance division: *pricing and asset-liability management of life reinsurance*
- since 2015: infinada Consulting GmbH, München: *model validation, management consulting*

tel: + 49 (151) 50512995
mail: stefan.jaschke@infinada.com
web: http://infinada.com